seit 1558

Complement Clauses and Complementation Systems: A Cross-Linguistic Study of Grammatical Organization

Dissertation zur Erlangung des akademischen Grades
eines Doctor philosophiae (Dr. phil.)

vorgelegt dem Rat der Philosophischen Fakultät
der Friedrich-Schiller-Universität Jena

von

Karsten Schmidtke-Bode, M.A.

geb. am 26.06.1981 in Eisenach

Gutachter:

1. Prof. Dr. Holger Diessel (Friedrich-Schiller-Universität Jena)

2. Prof. Dr. Volker Gast (Friedrich-Schiller-Universität Jena)

3. Prof. Dr. Martin Haspelmath (MPI für Evolutionäre Anthropologie Leipzig)

Tag der mündlichen Prüfung: 16.12.2014

Contents

Abbreviations and notational conventions

For each example sentence from a language other than English, an idiomatic translation is provided alongside a literal (interlinear 'morpheme-by-morpheme') one. The interlinear abbreviations given below largely conform to the *Leipzig Glossing Rules,* but also contain a number of modifications and particularly additions. I would like to thank two student assistants, Antonia Hülsebeck and Julia Kossiski, for their help in glossing the examples and synchronizing interlinear abbreviations across the sample languages.

Abbreviations for interlinear morpheme translation

1, 2, 3, 4	1st, 2nd, 3rd, 4th person	COP	copula
I, II, III, IV, V	morphological classes	CORE	non-nominative core case/article (Tukang Besi)
A	transitive agent (agent-like argument of canonical transitive verb)	CR	current relevance
		CTP	contemporative mood (West Greenlandic)
ABL	ablative	DAT	dative
ABS	absolutive	DEB	debitive
ACC	accusative	DECL	declarative
ACT	action (nominalizer), active	DEF	definite
ADD	additive	DEM	demonstrative
ADESS	adessive	DEP	dependent
AF	A-form (inflectional form in Urarina)	DESID	desiderative
		DETR	detransitivizer
AFF	affirmative	DF	D-form (inflectional form in Urarina)
AFR	aforementioned		
AGT	agent(ive)	DIR	directional, directive, directed; direct (evidential)
ALL	allative		
AN	animate	DIST	distal, distance
ANT	anterior	DO	direct object
ANTIP	antipassive	DS	different subject (switch-reference)
AOR	aorist		
APPAR	apparential (evidentiality)	DU	dual
APPL	applicative	DUR	durative
ART	article	EF	E-form (inflectional form in Urarina)
ASSOC	associative		
AUX	auxiliary	EFM	Euchee female (Yuchi)
BARE	bare pronoun (Lao)	EM	Euchee male (Yuchi)
BEN	benefactive	EMPH	emphatic
CAU	causal	ENC	encouragement (Yuchi)
CAUS	causative	ERG	ergative
CERT	certainty	EXCES	excessive
CLF	classifier	EXP	experiential (aspect)
CLIT	clitic	EXT	extent of action
CNTR	contrast(ive)	EZ	Ezafe (Persian)
COBL	oblique complementizing case (Kayardild)	F	feminine
		FACT	factitive
COM	comitative	FOC	focus
COMP	complementizer	FUT	future
COMPL	completive	GEN	genitive
CONN	connector, connective	GL	goal
CONT	continuous	HAB	habitual
CONV	converb		

HOD	hodiernal (past, future)
HON	honorific
HRS	hearsay (evidentiality)
HUM	human
IMAG	imaginative
IMP	imperative
IMPF	imperfect
IMPS	impersonal
INAN	inanimate
INCH	inchoative
INCL	inclusive
IND	indicative
INDEF	indefinite
INESS	inessive
INF	infinitive
INFL	inflectional complex
INFR	inferential
INS	instrumental
INT	intentional (mood, aspect)
INTR	intransitive, intransitivizer
INTS	intensifier
INVOL	involuntary
IO	indirect object
IPFV	imperfective
IRR	irrealis
ITER	iterative
L	low tone
LG	local gender (Motuna)
LGR	L-grade (aspectual class)
LINK	linker
LOC	locative
LOCUT	locutor marker (Awa Pit)
LOG	logophoric
M	masculine
MID	middle voice
MSD	masdar (gerund)
N	neuter
N-	non- (e.g. NSG non-singular, NPST non-past, NF non-feminine)
NEG	negative, negation
NGR	N-grade (aspectual class)
NOM	nominative
NOMFUT	nominal future
NMLZ	nominalizer
NR	near (tense or aspect)
NV	neutral version (Georgian)
OBJ	object
OBL	oblique
OPT	optative
P	transitive patient (patient-like argument of canonical transitive verb)
PASS	passive
PAUS	pausal

PCL	paucal
PL	plural
PNCT	punctual
POEL	postelative
POL	polite
POSS	possessive
PREVRB	preverb(al) marker
PRF	perfect
PRFV	perfective
PROG	progressive aspect
PROX	proximal, proximate
PRS	present
PSNV	presentative
PST	past
PTCP	participle
PURP	purpose, purposive
Q	question/interrogative
QUOT	quotative
RECP	reciprocal
REDUP	reduplication
REFL	reflexive
REL	relative clause marker
REM	remote (past, future)
REP	reportative, reported
RES	resultative
RL	realis
S	intransitive subject (single argument of canonical intransitive verb)
SBJ	subject
SBJV	subjunctive
SBST	substantivizer
SEQ	sequential, consecutive
SG	singular
SHRT	short
SIT	sitting position (Yuchi)
SPAC	spacer (Barasano)
SS	same subject (switch-reference)
STAT	stative
SUB	subordinate, subordinator
SUP	supine
TEMP	temporal
TERM	terminative
TH	thematic vowel or consonant
TNS	tense
TOD	today (past tense)
TOP	topic
TR	transitive, transitivizer
UG	undergoer
V	verb
VAL	valency increaser
VEN	ventive
VIS	visual (evidential)

Symbols

In both original language and interlinear morpheme translation:

x y	word boundary between x and y
x-y	morpheme boundary between x and y
x+y	x and y form a compound or a derivative stem
x=y	x and y are joined by clisis
$x_i...y_i$	x and y are coreferential elements

In original language only:

ø	null expression of meaning
[x]	x is a syntactic constituent (complement clauses are generally bracketed this way)

In interlinear morpheme translation only:

(x)	x is not overtly marked in the original (i.e. null expression of meaning)
x.y	x and y are grammatical (sub-)categories of one original language morpheme

Notational conventions

In example sentences, structures to be highlighted are printed in **boldface**. Language-particular lexemes or grammatical markers are generally given in *italics*.

The meaning of linguistic elements is usually given in 'inverted commas'. The names of linguistic categories are capitalized when reference is being made to a language-specific category (e.g. 'the German Infinitive'), but appear in non-capitalized form when the category is being used in a cross-linguistic sense (e.g. 'infinitives commonly derive from purposive action nominals').

Ungrammaticality is indicated by an *asterisk, and semantically odd sentences carry a ?question mark.

Spelling conventions

The present manuscript uses **British English spelling** throughout. Where variation in regard to the usage of <s> and <z> is found even within British English (e.g. *organise/organize*), the spelling with <z> is given preference.

1

Introduction

The present study is a cross-linguistic investigation into the grammar of complement clauses and the organization of complementation systems. It originated in the context of a larger research project on the typology of complex sentences, conducted from November 2007 onwards at the University of Jena. I would like to express my sincere gratitude to Holger Diessel for offering me to collaborate in this project over an extended period. It is this long-term support which has made it possible to collect the amount of data that underlies the present study, to see it in the larger context of the entire subordination systems of the languages to be investigated, and to thoroughly develop the theoretical ideas that will be brought to bear on the interpretation of the data.

The primary goal of the larger project is to provide a comprehensive typological analysis of the linear structure of complex sentences, i.e. of the ordering patterns that relative, object-complement and different types of adverbial clauses establish in relation to their respective attachment sites, and of the ways in which different ordering constellations correlate with aspects of the morphosyntactic and functional profile of the clause combination in question. The current dissertation is specifically concerned with complement clauses, and while aspects of linear order, morphosyntactic structure and grammatical function will also be central to the investigation, it represents an autonomous contribution in its own right. As will be explained shortly, it pursues markedly different goals, focuses on phenomena and data that are extraneous to the larger project, and approaches them in methodological ways that have been developed entirely independently. Consequently, I alone am responsible for how the data were selected, coded and analysed, and none of the people affiliated with the larger project is to be blamed for any errors or misguided interpretations.

The study of complement clauses is, of course, by no means new territory in linguistics, neither in philological nor in theoretical or cross-linguistic perspective. In fact, complementation is probably one of the best-studied syntactic phenomena of all (cf. Horie 2001 for a succinct overview of the typological research). As a result of this prominence, modern descriptions of hitherto undocumented languages also typically provide extensive discussion of complement clauses, so that there is now a relatively

large body of data available from genetically and geographically diverse languages (though still with marked biases towards specific families and areas, as will be discussed in Chapter 3). The present dissertation draws on this body of data in order to provide a thoroughly empirical account of the grammatical structure of complement clauses and complementation systems in the world's languages. The latter phenomenon – the organization of complementation systems – is, in fact, a noticeably understudied area in the otherwise well-developed typology of complementation. Much typological research has focused on the interaction of form and meaning in complement clauses (e.g. Givón 1980, Ransom 1986, Cristofaro 2003), but it is especially the work of R.M.W. Dixon (1995, 2006a) that has drawn attention to differences in how languages syntacticize complement relations and build different syntactic types of complementation system. Dixon's approach will be outlined properly in Chapter 2, but let me mention at this point that a large part of it has to do with the integration of complementation patterns into the argument-structural organization of individual languages, i.e. whether or not the complement functions as a genuine syntactic argument of a matrix clause, and which specific argument roles are typically covered by complement clauses when they propagate through the linguistic system over time. Indeed, much of the typological research on complementation is concerned with complements in object function (in the broadest comparative sense of the term 'object'), while no systematic investigation has been put forward of complements in other syntactic functions, notably as clausal subject. In great measure, the goal of the present work is to make inroads into this territory. It is not, however, a typology of subject clauses in a narrow sense, but casts a wider net that always keeps an eye on the development and organization of complementation systems, i.e. on similarities and differences of the structures being used for subject and object complementation, on the diachronic processes that drive the emergence of subject and object clauses, and on the ways in which the resulting complementation systems are to be characterized through Dixon's (1995, 2006a) syntactic lens. This perspective is reflected by the subtitle of the thesis, i.e. "a cross-linguistic study in grammatical organization".

More precisely, I am going to present four interconnected empirical studies of the grammatical organization of complementation, based on data from 100 languages. The first study, in Chapter 4, presents a comprehensive analysis of the internal structure of complement clauses from a cross-linguistic perspective. This chapter is foundational for the remaining ones in that it develops a catalogue of variables for measuring the degree to which complements approximate the grammatical properties of typical syntactic arguments, i.e. NPs, so that each complementation pattern in the data can be located on a scale of 'desententialization' in Lehmann's (1988) terms. The resulting metric will be used in later chapters to examine correlations between the internal structure of complement clauses and various aspects of their external syntax and function. In addition, the chapter will consider the central cross-linguistic claims on the internal properties of complement clauses made in survey works on complementation, notably in Noonan's (1985|2007) renowned overview article. Noonan presents a systematic account of the morphological types of complements commonly encountered (e.g. nominalizations or 'infinitives'), of syntactic processes

such as raising and control, of TAM distinctions and several other parameters of variation. Chapter 4 of this dissertation provides a qualitative extension and a quantitative underpinning of Noonan's morphosyntactic survey. In a sense, it is a fully empirical 'complement' to it, based on a principled sample of languages.

The second study, presented in Chapter 5, is concerned with an important external property of complementation patterns, namely their syntagmatic relation (or relative position) to the matrix. This phenomenon is investigated exclusively in regard to subject clauses, which are not in the purview of the larger project introduced above. The position of complement clauses in general, and of subject clauses in particular, has not yet been submitted to an in-depth typological analysis, although various proposals for preferred positional patterns and especially also their explanations can be found in the literature. These partly date back several decades, so the time is ripe for a reconsideration of the issue that is not only empirically well-founded, but also takes developments in the functional and usage-based literature into account for explaining the cross-linguistic situation. The result will be a fairly differentiated picture that includes considerations of information structure, online processing, diachronic evolution, forces of entrenchment and analogical extension, and the organization of the grammatical system. In developing this picture, I will also problematize the cross-linguistic applicability of the notion of 'extraposition' (which is commonly found in connection with complement clauses), and discuss the analytical challenges that are presented by subject clauses that are removed from the canonical subject position. These challenges become particularly significant if the Dixonian framework for analysing complementation systems is adopted, as the subject clauses in question often cease to be syntactic arguments of the matrix clause and would hence need to be treated somewhat differently in Dixon's account. Finally, I will also take a closer look at the place-holding morphemes that can sometimes be found as anticipatory elements to 'extraposed' subject clauses. While their cross-linguistic distribution is not particularly surprising, some interesting observations can be added from a functional, and particularly diachronic, perspective.

The studies in Chapters 6 and 7 are devoted to what has been called the 'selectional' relations of complement clauses and their matrix predicates. It is well-known that any given type of complementation structure co-occurs with a certain array of complement-taking predicates, and that these fall into more or less coherent classes, such as 'desiderative', 'phasal' or 'perception' verbs. Much typological work has been conducted on these co-occurrence patterns of complements and their matrix predicates, and the dissertation seeks to expand this research by adding new perspectives on the issue. First, I am going to define the predicate classes in such a way that they incorporate not only a semantic but also a syntactic dimension, distinguishing between subject- and object-clause-taking environments. In this way, it can be investigated how subject clauses fit into previously found typological generalizations on the predicate classes (such as Givón's (1980) famous 'binding scale' of complementation). Second, by carefully recording the co-occurrence relations between all complements and predicate classes in the sample, it becomes possible to analyse distributional similarities between the various predicate classes on the one

hand and between the individual complementation patterns on the other. It can be shown, for example, that there are marked asymmetries in how productively certain groups of matrix predicates co-occur with complement clauses across the world's languages: Classes associated with object complementation are considerably more productive than those involving subject clauses; and within both subject- and object-taking predicate classes, there are again principled differences in how often certain meanings (e.g. perception, propositional attitude, etc.) tend to be encoded by complementation or rather prefer alternative means of morphosyntactic expression. These differences will be described and motivated in detail in Chapter 6. That chapter is also the place to probe the more specific syntactic functions of complement clauses in the sample languages. For comparative purposes, complements can be said to occur as 'subjects' or 'objects', but how widely are they actually found, for example, as the direct object of a transitive verb? How commonly do we find complements that show the same morphosyntactic treatment as NP subjects in the respective language, and what kinds 'non-canonical' coding need to be recognized? In other words, I am going to examine (selected aspects of) how complement clauses are integrated with the argument-structural organization of languages.

The study in Chapter 7, finally, approaches the distributional profile of complement clauses from yet another perspective. It takes the synchronic co-occurrence patterns of complements and various predicate classes as the product of the gradual lexical diffusion, or analogical extension, of a complement from a historical starting point. One may ask, therefore, if the complements with similar distributional profiles in my sample are likely to have been derived from comparable historical sources. Given that only a fraction of the sample languages have detailed historical records of the diachronic development of complementation patterns, this question often needs to be approached in an indirect manner, but it will be seen that the results are very suggestive and that there are, indeed, intimate connections between certain 'diachronic types' of complement clause and pathways of lexical diffusion. This is an 'organizational' issue of complementation par excellence, as the diffusion patterns determine how each complement contributes to the structure and the grammatical ecology of the respective complementation system. At the end of Chapter 7, we will thus also have at least a brief look at typological differences in the ecological organization of complementation systems and, more generally, at how well developed complementation is as a grammatical process across the sample languages.

The empirical studies of the dissertation are preceded by two foundational chapters on the typological analysis of complementation. In Chapter 2, I gradually develop a suitable 'comparative concept' (Haspelmath 2010) of complement clauses for the present study and, in doing so, I discuss how the present approach integrates with (or differs from) previous typological approaches to complementation, argument structure, clause combinations, predicate classes, etc. These preliminaries will be vital in order to delimit the domain of investigation, but also to establish a conceptual and terminological framework for the entire study. In Chapter 3, I will outline some methodological principles that have guided the sampling, selection and coding of the data being used in this dissertation.

This brief overview of the structure of the dissertation will have made it apparent that the study does not pursue a single hypothesis or a straightforward line of argumentation for a specific theoretical point. It rather intends to make thoroughly empirical contributions to various (interconnected) aspects in the typology of complementation. Therefore, the overall aim of the dissertation is actually a very modest one. The individual topics to be covered make no claim to originality, but I hope that the methodological ways of approaching them, and at least some of the interpretations and conclusions, will be able to do so. Methodologically, a central characteristic of the present work is the application of explorative, data-mining techniques that are able to measure and visualize cross-linguistic similarities between complementation patterns, between complement-taking predicate classes and, finally, complementation systems. In this way, new perspectives on the cross-linguistic 'space' of complementation can be developed, in a profoundly data-driven manner.

In regard to the interpretative dimension, the patterns in the data will be investigated from the vantage point of usage-based models of language (cf. Kemmer and Barlow 2000, Diessel 2011 for surveys of this approach). This commitment is laid out in detail in two previous publications, i.e. Schmidtke-Bode 2009 (§2.2) and 2012, which is why I refrain from including a similar theoretical section in this introduction. What needs to be briefly commented on instead is how the processes of 'grammatical organization' that will figure prominently throughout the dissertation are conceptualized from a usage-based perspective. In alternative (i.e. formalist or nativist) accounts, grammatical regularities and system-building processes are typically attributed to the existence of 'syntactic principles', algorithmic mental 'rules' and similar constructs, which are also often taken to be specific to the domain of language. In usage-based models of language, by contrast, they are the product of the application domain-general processes (cf., e.g., Langacker 2000, Tomasello 2003, Dąbrowska 2004). The most important ones that we shall encounter in the following chapters are categorization and analogy.

Categorization involves the accommodation of the linguistic units perceived in a given utterance with one's stored linguistic representations (cf. Bybee 2010: 7). It is commonly assumed in usage-based accounts that these representations are 'exemplar-based', i.e. that they come in the form of clusters of experienced tokens that have been judged to be similar to each other (cf. Gahl and Yu 2006 for a brief introduction). If individual instances of language use are thus registered and stored, it follows that tokens with a high usage frequency have highly entrenched and hence accessible memory traces, and act as powerful 'attractors' for the categorization of newly encountered items (cf. Abbot-Smith and Tomasello 2006, among many others). Furthermore, it has been shown that highly entrenched exemplars tend to be diachronically inert, i.e. they are often retained even when other members of the category are being recategorized in language change (e.g. by succumbing to paradigm levelling, cf. Bybee 1985). These and related insights into categorization will also play a role in the present study, and sometimes in perhaps unexpected contexts (cf. e.g. Chapter 5).

An important point in usage-based linguistics is that certain tokens of an established category (especially of complex constructions) may show formal, semantic or contextual properties that match the attractors of a different category. In other words, such tokens potentially allow for recategorization, such that, for example, an established syntactic structure comes to be associated with a different functional category. This process is known as 'reanalysis' (cf. e.g. Croft 2000: Ch. 5), and it will concern us in great detail in Chapter 7, when we examine the diachronic sources of complementation. Processes of reanalyses fully assert themselves when another cognitive force is at work, namely analogical extension to novel contexts (Hopper and Traugott 2003: §3.5, Bybee 2010: 8). Analogy in this sense thus leads to the gradual implementation of a new form-function mapping (such as a new complementation pattern); it is a system-building process in the best sense of the term, and ultimately responsible for the creativity and 'rule'-based character of human language. In the present study, analogical extension will, of course, figure prominently in the context of lexical diffusion processes of complement clauses (Chapter 7), but forces of analogy in a more general sense can also be seen at work in many other places of this dissertation.

These selected usage-based processes are just the tip of the iceberg of the complex array of mechanisms and arenas that determine the organization of linguistic systems at large (cf. Beckner et al. 2009 for a more comprehensive overview). They have been introduced very briefly at this point to prepare the reader for the ways in which certain empirical findings will be framed in this study. With this general orientation in mind, we can now turn to the phenomenon of complementation as such and to the foundations of its typological study. This will be the topic of the following chapter.

2

The phenomenon of complementation

2.1 Introduction

The present study concerns itself with biclausal syntactic constructions in which the predicate of one clause "entails reference to another proposition or state of affairs" (Cristofaro 2003: 95), expressed in a second clause. For example, the English predicate *want* denotes a mental activity that is inherently directed at[1], and hence entails reference to, another state of affairs:

(1)   *I want* [*to buy a new car*].

This linguistic phenomenon goes by the name of **complementation**, and perhaps the most common way of framing the issue is by saying that "a predication" comes to function as "an argument of a predicate" (Noonan 2007: 52). When this situation is expressed in individual languages, the resulting utterances and their constituent parts will be described in this study by recourse to the following grammatical terminology: The example in (1) may be characterized in grammatical terms as a **complement sentence** (following Cristofaro 2008) involving a **complement-taking predicate** *want* (henceforth **CTP**, following Noonan 2007) and a **complement clause**, the bracketed part of (1). When the complement clause is removed from the sentence in (1), the unit that remains will be referred to as the **matrix** or **matrix clause**[2]. Finally, the whole sentence as it actually occurs in language use is a specific instance of a **complementation construction** (in the sense of Construction Grammar (e.g. Goldberg 1995, 2006; Croft 2001)). For reasons to be discussed later on (§2.5), the term 'complementation construction' is thus deliberately reserved for the abstract schema of which the *entire* sentence (1) is an instantiation, rather than for the complement clause alone (as is normally done in the literature).

---

[1] The expression 'directed at' is taken from language philosophy, in particular Searle (1983), who refers to the property of mental states and events to be directed at other propositions or states of affairs as 'intentionality'.

[2] In (1), the sequence *I want* is not, strictly speaking, a clause of its own and hence the more neutral term 'matrix' would be used in such instances. There are cases, however, in which the matrix does consist of an entire clause, and in such cases, we will speak of the 'matrix clause'.

Cristofaro's and Noonan's definitions of complementation capture the phenomenon in semantic terms, i.e. they provide a basis for an onomasiological study of the morphosyntactic means that individual languages employ for expressing reference to another predication. This perspective is also embodied in Deutscher's (2000) notion of '**FDC**' – the **functional domain of complementation** –, which he uses to refer to the totality of grammatical techniques that, at any given point in time, serve the purpose of (semantic) complementation within and across languages. As Deutscher notes, these techniques can be quite heterogeneous. For example, all of the sentences in (2) may be said to be part of the FDC, although only (2b) is commonly recognized as a 'complement clause' in grammatical descriptions of English:

(2) a. *I can't understand* [*John's immediate refusal of the offer*].
   b. *I can't understand* [*that John refused the offer immediately*].
   c. *John refused the offer immediately, and I can't understand it.*

In the sentences in (2), the grammatical units that express John's refusal of the offer differ not only in their internal morphosyntactic composition (assuming the structure of a noun phrase in (2a), that of a subordinate clause in (2b) and of an independent clause in (2c)), but also in the way they are syntactically integrated with the rest of the sentence (embedded as a constituent of the sentence in (2a) and (2b), but part of a coordinate clause combination in (2c)). Taken together, it is only in relation to (2b) that one could argue that reference to another predication has been **syntacticized** into a truly **biclausal** construction. In order to distinguish such complement clauses 'proper' from functionally related constructions, and to draw insightful generalizations about the organization of the FDC across the world's languages, typologists are required to delimit the core of the FDC in precise and cross-linguistically commensurable ways. The most explicitly worked-out proposal for such a delimitation that I am aware of has been put forward by R.M.W. Dixon.

Initiated in Dixon (1995) and fully articulated in Dixon (2006a), his approach draws a sharp distinction between **complement clauses** (proper) and **complementation strategies**. In order for a grammatical constituent to be recognized as a complement clause, it needs to meet the following four criteria:

(3) Grammatical properties of complement clauses (adapted from Dixon 2006a: 15)

  i. A complement clause functions as a syntactic argument of a higher clause. [This higher or superordinate clause was called a 'complement sentence' above, and a more widely used alternative term for it is **main clause**.]

  ii. A complement clause has the internal structure of a clause. Especially its arguments, if not omitted by a grammatical rule specifically associated with the complement clause in question, should be marked in the same way as in main clauses, and have much the same grammatical properties.

  iii. A complement clause will always describe a proposition. This can be a fact, an activity or a potential state.

  iv. In every language that has complement clauses, they function as a core argument for verbs with meanings such as 'see', 'know', 'believe', and 'like', for

'tell' if there is an indirect speech construction, and for 'want' if this concept is
realized as a lexical verb.

Each of these criteria for complement-clause status deserves careful attention. On the
one hand, this is because the above descriptions need to be fleshed out to be fully
appreciated, and they also need to be made more precise so that they will be specifically
tailored to the goals of the present study. On the other hand, each of the four aspects of
Dixon's definition has a history in typological research and thus comes with a large
amount of previous literature (e.g. on argument structure, clause combinations,
semantic classes of verbs, etc.). Consequently, a second aim of the present chapter is to
use the above criteria not only to delimit the scope of the investigation, but also to
integrate it systematically with the conceptual and terminological conventions in the
relevant areas of cross-linguistic research. To this end, each of the following sections in
the present chapter will elaborate on one of Dixon's criteria above, and on the concepts
that it makes reference to. In §2.2, we will concern ourselves with the notion of
syntactic arguments and the integration of complement clauses into the typology of
argument structures and grammatical relations. §2.3 is devoted to the notion of 'clause'
in the context of complex sentences, clarifying what counts as a clause on
morphosyntactic grounds and which structures fall outside the scope of the study. In
§2.4, I will scrutinize Dixon's third criterion from above to exclude further
constructions, but this time for semantic rather than purely syntactic reasons. And in
§2.5, I will take issue with Dixon's last criterion and carefully establish our own set of
complement-taking predicates.

Throughout this and the remaining chapters of the dissertation, it will be useful to
have a cover term that subsumes both complement clauses and complementation
strategies as grammatical techniques for expressing the FDC. I will use the term
'**complementation pattern**' in this neutral, non-committed way. Adding this notion to
the grammatical terminology from above, we may thus say that a given complement
sentence consists of a main clause housing a CTP, and a complementation pattern of a
specific morphosyntactic type.

2.2 Argument status

Dixon's requirement that complement clauses function as a syntactic argument of
another clause is a particular instance of the concept of '**embedding**' (Foley and
VanValin 1984, Matthiesen and Thompson 1988, Lehmann 1988), i.e. the syntactic
relationship that holds when a clause forms (part of) a constituent in the main clause.
This situation obtains in (4) below from Korean:

(4)   Korean (isolate: North Korea, South Korea; Sohn 1994: 55)

   Na-nun [*apeci-ka*   *o-si-ess-um-ul*]   *al-ass-ta.*
   1SG-TOP   father-NOM come-HON-PST-NMLZ-ACC   know-PST-DECL
   'I knew that father came.'

In this example, the complement clause directly replaces a direct object (signalled by
the accusative case marker) and is thus embedded as an argument of the main clause.
In complementation, embedding contrasts with the left- or right-**adjunction** of a

clause to a main clause that is fully saturated in its argument structure. Consider the following example from To'aba'ita:

(5)   To'aba'ita (Austronesian, Eastern Malayo-Polynesian, Oceanic: Solomon Islands; Lichtenberk 2008: 984)

Meresina   qeri   qe   ade-a   wane   [ka   qiiqida qasia naqa].
medicine   that 3SG.NFUT   do-3OBJ   man   3SG.SEQ   sweat   INTS   INTS
'The medicine made the man sweat a lot.' (Lit.: 'The medicine did the man, he sweated a lot.')

In this sentence, the matrix verb *ade* 'do' has a causative interpretation and thus entails reference to another state of affairs; we would thus expect it to take a clause as its complement. Syntactically, however, the clause in brackets is adjoined to a structurally saturated matrix clause that could stand alone as an independent sentence, since *ade* already takes the NP *wane* 'man' as its direct object. As Lichtenberk (2008: 984–985) comments, "it is the causee phrase and not the complement clause that is the object of the causative verb"; therefore, the clause in brackets is not embedded as a syntactic argument of the matrix predicate. In Dixon's framework, then, it does not qualify as a complement clause, but, being functionally equivalent to such a construction, can be considered a complementation strategy. More specific terms that have been used for this kind of complementation strategy include **apposition**, **parataxis** and **juxtaposition** (Dixon 2006a: 38).

In the present study, I follow Dixon's decision here: The complementation patterns considered for the analysis (cf. Chapter 3 for their sampling) will be classified into complement clauses and complementation strategies, and Dixon's criterion of argument status will be one parameter determining the classification. The resulting division harbours the potential of discovering different syntactic types of complementation systems, viz. those that grammaticalize complementation into a syntactic relationship of embedding, and those that rely exclusively, or partially, on more paratactic ways of expressing the same concepts. However, stating that a clause functions as a syntactic argument of another clause actually entails two much more fundamental questions: First, which kinds of syntactic argument can we recognize cross-linguistically, and what are the comparative labels we shall give to them in the present study ('subject', 'object', etc.)? This question makes it necessary to expose some basic assumptions about argument structure from a comparative point of view, thereby situating complementation in what is called 'grammatical-relations typology'. This will be done in §2.2.1. Second, on what grounds can we claim that an NP or a complement clause has 'argument' status to begin with? This question will be dealt with in §2.2.2: It contains a brief problematization of the very notion of 'syntactic argument', which will be taken for granted up to that point but deserves some comments in its own right.

2.2.1 Complement clauses and argument-structure typology

As, for example, the work by Bickel (2011) and Witzlack-Makarevich (2010) demonstrates, the typological study of grammatical relations has changed considerably over the last decades. In the wake of these developments, the usage of widely applied

comparative concepts, such as 'intransitive clause', 'transitive subject', 'S, A, P', etc. has diversified even among similar-minded typologists. Haspelmath (2011a) is determined to outline these terminological changes and inconsistencies, and reminds us that future studies in this area must "feel compelled to provide a precise definition" (ibid.: 537) of the terminology being used to compare argument-structural phenomena across languages. This is what the present section intends to do. Although the discussion of these definitional issues may seem unnecessarily long-winded and cumbersome, it is necessary in order to arrive at a grid of comparative concepts that is specifically geared to the present purposes.

The present study will draw on the following model of argument-structural patterns developed by Croft (2001) (Fig. 1):
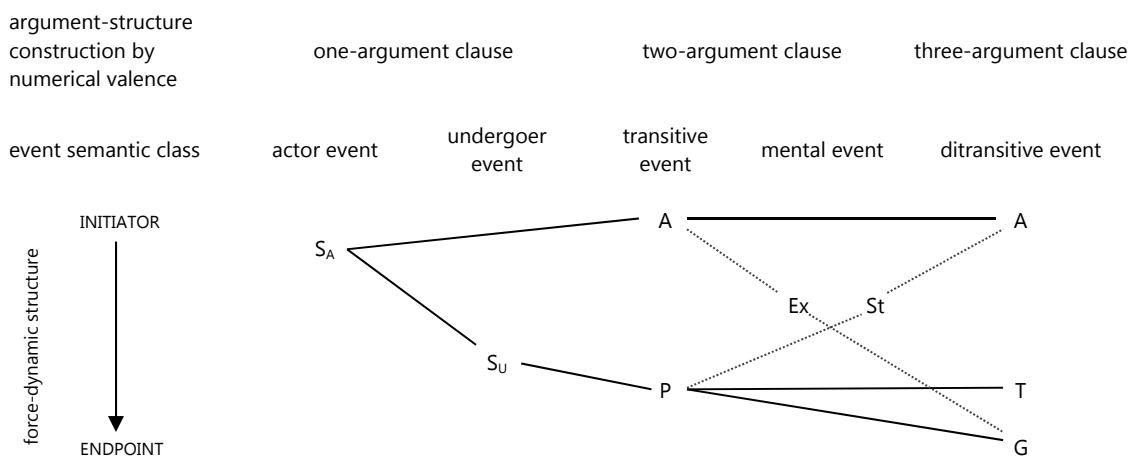
Figure 1. Conceptual space of generalized semantic roles of syntactic arguments (adapted from Croft 2001: 164, with some additions)[3]

The starting point of virtually all typological research into grammatical relations is that clauses can be characterized in terms of the number of syntactic arguments they contain. As we shall see in §2.2.2 below, this traditional assumption of **numerical valence** is not without problems, but we shall employ it to make a basic contrast between one-argument, two-argument and three-argument clauses. The core of Croft's model then aims to relate those clause patterns to different semantic classes of events. The latter, in turn, are characterized by their **force-dynamic structure**: Specifically, events can be described in terms of their transmission of force from an initiating participant to an endpoint. In other words, events with multiple participants define a causal chain on which the participants claim different ranks relative to each other (cf. also Talmy 2000: ch.7, Langacker 1991: 283). This is the cognitive-linguistic formulation of the widely cited idea in typology that participants can be characterized as 'agent'-like or 'patient'-like relative to one another. In clauses with a single argument

---

[3] In the following, I will concentrate on the types of events, and the corresponding linguistic labels for the arguments involved, that Croft's model distinguishes. For the present purposes, we will neglect the various lines that connect them and their underlying logic, as these are not needed for the exposition.

(labelled **S**), the participant expressed by this argument may also be 'agentive' or 'patientive', resulting in what Croft calls 'actor events' and 'undergoer events', respectively.

Moving rightwards in the model, we enter the realm of two-argument clauses, and it is here that an interesting split in the force-dynamic structure occurs. On the one hand, we have two-participant events that fully conform to the prototypical causal chain: they constitute "an effective volitional discrete action performed by a controlling agent and actually affecting a well individuated patient" (Lazard 2002: 152). Croft calls them transitive events, and they give rise across languages to what Andrews (2007a: 138) terms "primary transitive verbs" like *kill, break, push,* etc. Accordingly, the argument-structural construction these verbs normally appear in is known as a '**transitive clause**' or the 'major biactant construction' (Lazard 2002: 153) of the language in question. Furthermore, as can be seen in Fig. 1, the syntactic functions embodied by transitive constructions are called **A** and **P**, mnemonic for the more agent-like and more patient-like argument, respectively.

On the other hand, however, there are also two-participant events that deviate from the transitive prototype. Semantically, they chiefly comprise what Croft refers to as 'mental events'. These involve the generalized participant roles of stimulus and experiencer, and in the literature, the predicates coding such events are variously known as 'experiencer verbs', 'psychological verbs', or simply 'psych verbs'. This class of verbs turns out to be particularly important for complementation, as many of the typical complement-taking predicates precisely denote mental states ('see', 'know', 'think', 'believe', 'want', 'need', 'seem', etc.). In the literature, they have sometimes also been referred to as 'low-transitivity' predicates in Hopper and Thompson's (1980) sense, owing to the fact that the transmission of force in mental events is cyclic rather than unidirectional or asymmetric: "the experiencer directs her mental attention to the stimulus, and the stimulus causes a change of (mental) state in the experiencer." (Croft 2012: 233) It is for this reason that the semantic roles 'St' and 'Ex' show up in the middle of the force-dynamic continuum in Croft's model (i.e. in between typical initiators and endpoints), and that they do not differ on this dimension relative to each other. As Kutscher (2009: 43) puts it,

> "Psychverben weisen einheitlich eine bidirektionale Kausalrelation auf, die [...] inner- und übereinzelsprachliche Konstruktionsvielfalt nach sich zieht."
> [Psych verbs are uniformly characterized by a bidirectional causal relationship, which leads to structural diversity in their coding across languages.[4]]

Indeed, it is an empirical question whether speakers of a given language subsume mental predicates denoting 'see', 'know' or 'want' under the major two-argument construction, thus investing them with fully transitive coding, or whether they choose a so-called '**non-canonical construction**' (Onishi 2001) for this purpose. A variety of such deviant coding types is illustrated in the following examples:

---

[4] Essentially the same conclusion can be found in many other comparative studies of experiencer verbs, such as Bossong (1998), Næss (2007: ch.8) or Verhoeven (2010).

(6)   a.   Turkish (Altaic, Turkic: Turkey; Göksel and Kerslake 2005: 128)

*Herkes   piyanist-**e**   bayıl-dı.*
everyone   pianist-**DAT**   adore-PRFV
'Everyone adored the pianist.'

b.   Djaru (Pama-Nyungan: Australia; Tsunoda 1981: 113)

*Jambagina   ŋa=**ɲanda**   juwa maɲ-an   guɲar-**a.***
child.ABS   CLITIC.HOST=**3SG.LOC**   fear-PRS   dog-**LOC**
'A child is afraid of a dog.'

c.   Slave (Na-Dene, Athapaskan: Canada; Rice 1989: 1224)

Kuyenį̣wę.
3:SBJ.wants.3PL:**OBL**
'He wants them.'

d.   Matsés (Panoan: Brazil; Fleck 2003: 1035)

*Natia   mani-ø   bun-e-**bi.***
strongly   plantain-ABS   want-NPST-1.**S**
'I strongly desire plantains.'

All of these examples differ from the canonical transitive construction in the respective language. A deviation in the non-subject argument is commonly referred to by the term **oblique object** (as opposed to direct object). This argument receives Dative coding in Turkish and Locative coding in Djaru, respectively. In Slave, it is indexed by "an oblique object pronoun" on the verb (Rice 1989: 1224). What we see in (6d) from Matsés is that the object retains the canonical Absolutive form but the subject argument does not appear as a canonical A argument. Instead, the experiencer in (6d) would appear in the Absolutive case if it were a full NP, and it is also indexed on the verb like the S argument of a typical one-argument clause. More generally, clauses like (6a–d) above have traditionally been referred to as **intransitive clauses**, and the non-subject argument of these clauses has been given the label '**E**' by Dixon (1995, 2006a), mnemonic for 'extension'.[5]

However, it has to be noted that even the recognition of a fourth core argument in addition to S, A and P does not exhaust the possible space of argument-structural patterns across languages. For example, as Haspelmath (2011a) shows, Lezgian has a non-canonical two-argument construction involving a Dative NP and a Postelative NP:

(7)   Lezgian (Nakh-Daghestanian: Azerbaijan, Russia; Haspelmath 1993: 281)

Šarwili.di-z   ada-qʰaj   kič'e   x̂a-na-č.
Šarwili-DAT   he-POEL   afraid   be-AOR-NEG
'Šarwili was not afraid of him.'

---

[5] It is unclear whether Dixon himself would subsume all non-subject arguments in (6) above under his label 'E'. In Dixon (2006a), he conceives of S, A, P and E as 'core arguments', suggesting that they collectively contrast with peripheral arguments or adjuncts, and that E arguments are "typically marked by the dative" (ibid.: 8). But the actual extension of the category E is never made explicit.

As is noted by Haspelmath, none of the so-called 'core arguments' commonly assumed in typology, including Dixon's E, would account for this pattern.[6]

Similarly, there are, of course, also non-canonical versions of one-argument clauses. A famous case in point is the Indic language Sinhala. In this language, the single argument of a monovalent clause pattern is normally accorded the Nominative case, but depending on various semantic factors, it may also surface with Dative or even Accusative marking, as in the following example discussed in Næss (2007: 207):

(8)　　Sinhala (Indo-European, Indic: India)

　　　a.　*Miniha　　duwənəwa.*
　　　　　man.NOM　run.PRS
　　　　　'The man runs.'

　　　b.　*Minihatə　　diwenəwa.*
　　　　　man.DAT　　run.INVOL.PRS
　　　　　'The man runs (involuntarily).'

Again, it has not been customary in typology to call the Dative NP in (8b) an 'S' argument, since this label is reserved for the canonical cases of one-argument clauses. In fact, however, it is well-known that one-argument clauses are anything but semantically homogeneous and that it is actually very common for languages *not* to treat all instances of this clause pattern in the same way (cf. Dryer 2007: 259ff. for an overview). For example, we often find a grammatical distinction between stative and non-stative monovalent predicates, with the latter being canonical in morphosyntactic behaviour (e.g. aligned with the A argument of transitive clauses) and the former deviating in several ways to form a morphosyntactic class of their own. A variation on this pattern is what has been called 'split intransitivity', in which agentive S-arguments (S_A in Fig. 1) are systematically aligned with A, while patientive or undergoer arguments (S_U) are aligned with P. In such situations, we have two canonical patterns of one-argument clauses and (at least in some cases) no non-canonical class that differs in coding from the core arguments.

However, in view of the fact that non-canonical marking generally needs to be taken into account in grammatical-relations typology, and for complement-taking predicates in particular, I share the need felt by some researchers to broaden the traditional labels of S, A and P to cover canonical and non-canonical constructions alike (rather than introducing ever new labels for all kinds of non-canonical pattern, or discarding the latter from the scope of typological enquiry). Representatives of this approach are Bickel (2011) and Witzlack-Makarevich (2010). In the present study, I will take a similar, though not entirely identical approach. Specifically,

---

[6] The point of Haspelmath's (2011a) paper is actually to argue that this is not a problematic state of affairs, since, in his opinion, the most fruitful generalizations in grammatical-relations typology can be captured in terms of S, A and P defined with respect to prototypical two-participant and one-participant events. However, he concedes that for some typological questions, the properties of "non-prototypical two-argument verbs will eventually have to be drawn into the comparison" (ibid.: 551), and complementation is precisely an area of cross-linguistic research that makes it necessary to pay closer attention to non-canonical argument structures.

- I subsume the patient-like arguments of *all* two-argument clauses under the label 'P'. In non-canonical constructions like (6a) above, the dative NP *piyaniste* would thus be a P argument. The specific criteria for deciding on which argument of a non-canonical clause is construed as the more patientive one will be discussed below. In accordance with the classic typological terminology, I will refer to clauses containing *canonical* A and P arguments as transitive clauses.

- I subsume the relatively more agent-like arguments of *all* two-argument clauses under the label 'A'. In non-canonical constructions like (6b) above, the Absolutive NP *jambagina* would thus be an A argument, rather than an S in most typological approaches. Again, the specific criteria for deciding on the agentive construal of arguments in non-canonical clauses will be discussed below.

- I subsume the single arguments of *all* one-argument clauses under the label 'S'. In non-canonical constructions like (8b) above, the dative NP *minhatə* would thus be an S argument, despite its non-canonical coding.

- I reserve the terms 'transitive' and 'intransitive' to particular *coding patterns* of one-argument and two-argument clauses, in keeping with classic typological terminology. In my approach, then, these terms are not directly related to the labels S, A and P.

In other words, my usage of the labels S, A and P is primarily based on the numerical valence of clause patterns, and on a binary force-dynamic distinction in two-argument clauses. I differ from Witzlack-Makarevich (2010) in that I do not conceive of S, A and P as labels that apply to the semantic arguments of specific predicates, but to the syntactic arguments of certain types of clauses. This becomes apparent, for example, in the treatment of passive constructions. In keeping with traditional typological research, I conceive of passives derived from transitive clauses as one-argument clauses, with the agentive participant being demoted to adjunct status (e.g. *The window was broken (by John)*). Therefore, there is only a single syntactic argument to be recognized in such constructions, which thus receives the label 'S'. Witzlack-Makarevich, by contrast, proceeds from the predicate found in such constructions (e.g. *break*), which takes two semantic arguments (here: *John, the window*), and those can, of course, still occur in the passive clause. Accordingly, she applies the labels A and P not only to the active but also to the passive version of the clause above. In other words, her labels S, A and P are really to be seen as referring to generalized semantic roles of predicates which apply regardless of the particular syntactic frame in which the predicate in question occurs.

To sum up my terminological approach so far, let us have a look at Table 1. It cross-classifies my comparative usage of the labels S, A and P with the question of whether the corresponding clause pattern in which they occur receives canonical or non-canonical coding, based on the examples provided above:

Table 1. S, A and P in canonically and non-canonically marked clauses[7]

	S	A	P
canonical	*miniha* in (8a)	*nanun* in (4)	*wane* in (5)
non-canonical	*minihatə* in (8b)	*jambagina* in (6b)	*piyaniste* in (6a)

This grid can now be applied to complement clauses. When a complement clause replaces the single argument NP of a one-argument construction, it functions as S. In this function, the complement typically encodes a proposition or state of affairs that is appraised along a variety of different dimensions, such as by an epistemic, evidential or deontic judgement (*It seems/is likely/is good/was necessary/was reported* [*that Peter left his wife*].). As such, it complements the sense of a normally stative predicator in the main clause. Depending on whether such stative predicators appear in canonical monovalent clauses (as in English) or whether they constitute a class that deviates in its clause structure (e.g. by lacking indexation for the argument), they will be said to instantiate canonical-S and non-canonical-S clauses, respectively. (Examples will be given in the analytical chapters of the dissertation.)

When complement clauses appear in two-argument clauses, they commonly code the stimulus of a mental event. As was discussed above, these events are characterized by a bidirectional flow of energy between experiencer and stimulus, and "either [of them] may form the basis of the construal of the causal structure of the mental event" (Croft 2012: 233f.). Across languages, this results in three different patterns of lexicalization for mental predicates, which have implications on which argument is to be seen as 'more agent-like' (i.e. 'A') and which one as 'more patient-like' (i.e. 'P') for comparative purposes.

In the first lexicalization pattern, the predicate construes the stimulus as actively causing a mental (change of) state in the experiencer. In such cases, the stimulus outranks the experiencer on the force-dynamic scale, and, across languages, "the stimulus is consistently realized as [A] and the experiencer as [P]" (ibid.) in my use of those labels. In the literature, predicates of this kind have been called '**experiencer-object verbs**' (e.g. Belletti and Rizzi 1988, Pesetsky 1995), and such lexicalization patterns lead to complement clauses in 'A' function, as in the following cases from English:

(9)   [*That Pete shouts at his children*] *bothers/annoys/pleases/frightens me.*

It is now a question of morphosyntactic coding whether the form of experiencer and stimulus in such constellations corresponds to that of a canonical transitive clause or to that of a different, non-canonical, pattern. Compare the following examples from German (with glossing reduced to the significant contrasts):

---

[7] The table simplifies the situation somewhat because it disregards situations in which there is reason to assume two canonical patterns side by side (e.g. certain kinds of 'split intransitivity' or two equally important voice systems of two-argument clauses in many Austronesian languages, etc.).

(10) German (Indo-European, Germanic: Germany, Austria, Switzerland)

    a. *beunruhigen* 'worry' in the canonical (transitive) two-argument construction

        Die   *Nachricht*   *beunruhigt* *den*   *Jungen.*
        the.NOM news.NOM   worries   the.ACC boy.ACC
        'The news worries the boy.'

    b. *gefallen* 'please' in a non-canonical two-argument construction

        Die   *Bücher*   *gefallen*   *dem*   *Jungen.*
        the.PL.NOM book.PL.NOM   please.PL the.SG.DAT boy.SG.DAT
        'The books please the boy.'

In (10a), the mental verb *beunruhigen* 'worry' is cast in the transitive construction, with a Nominative A argument and an Accusative-marked P; by contrast, the mental verb *gefallen* 'please' in (10b) shows non-canonical coding, the stimulus appearing as the subject (receiving Nominative case marking and triggering indexation on the verb) and the experiencer as an oblique (Dative) argument. Accordingly, when the Nominative argument in (10a) and (10b) is replaced by a complement clause, we can speak of canonical and non-canonical A-clauses, respectively.

A second lexicalization pattern is found when the predicate profiles the experiencer as actively attending to the stimulus. In such cases, the experiencer outranks the stimulus on the force-dynamic scale, and, across languages, "the experiencer is consistently realized as [A] and the stimulus is realized as [P]" (Croft 2012: 234) in my use of those labels. In English, this '**experiencer-subject**' lexicalization pattern is found with mental activity predicates like *ponder* or *consider (a problem)* or *watch* (*an event*). If the stimulus is expressed by a complement clause, this clause thus assumes 'P' function. And again depending on the morphosyntactic kind of two-argument construction chosen for this purpose, the complement clause can be a canonical-P or non-canonical-P clause, respectively.[8]

The third and most important lexicalization pattern in our context is found when the predicate "construes the experiencer-stimulus relationship as a state, i.e. as force-dynamically neutral" (ibid.). Put differently, experiencer and stimulus do not differ substantially (if at all) in their position in the causal chain, so that there is no inherent ranking with regard to 'agent'- and 'patient'-like properties. As a result, the allocation of the two arguments to particular case frames differs enormously across languages, and it can sometimes be difficult to decide which argument is to be called 'A' and which one 'P'. The analysis is straightforward when the mental state is assimilated to the transitive construction, as in the following example from Dolakha Newar:

(11) Dolakha Newar (Sino-Tibetan, Tibeto-Burman, Bodic: Nepal; Genetti 2006: 145)

    Āpsin   [*ām jāl ta-en*   *ta-e*   *khã*]   *ma-si-u.*
    3PL.ERG that net put-PTCP put-NMLZ COMP NEG-know-3PST.HAB
    'They didn't know that the net had been put there.'

---

[8] Predicates of this kind are typically fewer in number than the members of other lexicalization patterns of mental predicates. Therefore, they will play only a minor role in the present study and I refrain from providing examples of the corresponding complement clauses here.

In this example, the experiencer carries the Ergative case marker typical of the transitive construction, and the complement clause appears in the object position between the subject and the main verb. Therefore, we are dealing with an 'experiencer-subject' pattern overall, and the complement qualifies as a canonical-P clause in our terminology. The same kind of construal can sometimes also be argued for when a non-canonical clause pattern is involved. In the Mongolic language Buryat, for instance, the mental state verb denoting 'regret' codes the experiencer NP in Nominative case and indexes it like canonical subjects, while the stimulus argument appears in the Dative case (rather than the Accusative that would be characteristic of a transitive clause). Thus in (12) below, the participial complement clause can be said to be a non-canonical P-clause in our classification.

(12) Buryat (Altaic, Mongolic: China, Mongolia, Russia; Skribnik 2013)
[*Tiime bolo-so-hon-**do**-ny*]   *xalagla-na-b.*
this   become-INTS-PTCP.PST-**DAT**-3SG   regret-PRS-1SG
'I regret that it happened like this.'

Analytical problems arise when languages distribute the properties that identify subjects and objects in the first place. Consider example (13) from Kannada:

(13) Kannada (Dravidian: India; Sridhar 1976: 584)
Nanage i vicara gottu.
1SG.DAT this fact.NOM know
'I know this fact.'

Croft (2012: 234) analyses this example as instantiating an 'experiencer-object' pattern and the stimulus as being the subject of the clause. Grammatical evidence for this position comes from the fact that the stimulus *i vicara* receives the case marking normally reserved for the agents of transitive clauses, i.e. the Nominative, and that the experiencer exhibits the Dative case, normally associated with oblique arguments. However, it is not uncommon in the Dravidian literature to speak of 'dative-subject' constructions (cf., e.g., Krishnamurti 2003: 425[9]), on the grounds that the experiencer argument in (13) seems to appear in the canonical subject position (SOV) and that it exhibits so-called 'behavioural properties' normally associated with subjects. In general, such behavioural properties include the ability to act as a controller of co-reference relationships (for implicit arguments in subordinate or coordinate clauses, or for reflexive pronouns in main clauses), as the addressee of imperatives, etc. (cf. Onishi 2001, Falk 2006: ch.1 for recent overviews). The reason why apparently oblique-marked experiencers receive such subject-like treatment on behavioural grounds is well-understood: Since experiencer arguments usually refer to a definite human participant (as in (13) above), they are often moved into a topical position in the sentence, and since "most human topics are subjects", the oblique NP "is gradually assimilated to subjects with respect to its morphosyntactic behaviour" (Haspelmath

---

[9] Krishnamurti's analysis thus reads: Sentences like (13) above "can be interpreted as a transitive sentence with the subject in the dative and the object in the nominative. In such sentences, the object is also a core argument, since non-mention of the object [...] produces an ungrammatical sentence." (Note that 'transitive' is being used as a synonym of 'two-argument clause' here and not in the typological sense established above.)

2001: 78; cf. also Bickel 2004: 77 and the papers in Serzant and Kulikov 2013). As has been shown in detail by Cole et al. (1980), this process tends to affect behavioural properties before coding properties; that is, the dative coding persists while the NP is gradually allowed to participate in processes like complement control, conjunction reduction, reflexivization, etc. Depending on how widely such syntactic operations define subjecthood in a given language, dative experiencers will be called more or less subject-like in constructions like (13) above. In other words, the synchronic situation may show a certain clash between coding and behavioural properties, which makes such patterns challenging to analyse: In examples like (13) above, should we go with the coding properties to say that 'this fact' is A while 'me' is P, or with the (possible) behavioural evidence to arrive at the opposite classification? Clearly, the problem extends from phrasal to clausal arguments, so that we face the same challenge with regard to complement clauses. As a case in point, we can turn to another Dravidian language, Malayalam, in which the finite complement clause marked by *ennə* occurs in a clause pattern that is entirely analogous to (13) above:

(14) Malayalam (Dravidian: India; Jayaseelan 2004: 232)

    En-ik'k'e [*Mary miDukki*   *aaNə* *ennə*]   *toonn-i.*

    1SG-DAT Mary clever.person be.PRS COMP   seem-PST

    'It seemed to me that Mary is clever.'

Jayaseelan shows that the behavioural properties commonly adduced for the alleged subject status of the experiencer fail (or are inconclusive). As a result, he follows Shibatani (1999) in his verdict that the notion of 'dative subject' is wrong-headed in Malayalam and that, instead, clauses like (14) should be treated as intransitive constructions with a Nominative subject. From this perspective, then, we would be dealing with a complement clause in 'subject' function; and under the assumption that "the dative NP is an oblique argument" (Jayaseelan 2004: 241) rather than an adjunct, the complement clause in (14) would be placed in our 'non-canonical A-clause' category.

    Ideally, then, we would fathom, for each of the non-canonical constructions in our sample, how strongly the experiencer has developed behavioural properties that can confirm its subject status. However, unless this discussion has been held for the language in question (e.g. Malayalam, Lezgian and a few others), this case-by-case investigation is beyond the scope of the present study. For this reason, I will have to stick to the more overt coding properties and the extent to which they support a particular analysis (similarly to Croft and his verdict on (13) above). This will normally involve the constellation of case marking, indexation and word order. Let us recall the example from Matsés we encountered earlier, repeated as (15a) here for convenience:

(15) Matsés (Panoan: Brazil; Fleck 2003: 1035)

    a. *Natia*   *mani-ø*   *bun-e-**bi**.*

        strongly plantain-ABS want-NPST-1.**S**

        'I strongly desire plantains.'

    b. *Debi-ø*   [*nid-te*]   *bun-e-k.*

        Davy-ABS go-COMP want-NPST-IND

'Davy wants to go.'

In Matsés, the predicate *bun* 'want' takes two arguments, which *both* appear in Absolutive case if they are coded by full NPs. The case marking thus perfectly reflects the symmetrical force-dynamic relationship of experiencer and stimulus in mental-state predicates. However, it is the experiencer that receives indexation on the verb and appears in the canonical subject position; therefore, we can be reasonably certain that the experiencer is best analysed as (non-canonical) A and the stimulus as P in our coding scheme. Analogously, the complement clause in (15b) will also be analysed as an 'object clause' in our terminology, as there is no compelling evidence that it functions as an S or A argument on morphosyntactic grounds. The overall relevance of these analytical details becomes apparent when we want to make generalizations about the syntactic functions that complement clauses can, or cannot, fulfil across the languages of our sample (an issue to be dealt with in Chapter 6). Clearly, we will want to make a difference between languages in which the occurrence of 'subject clauses' is limited to non-canonical constructions, and those which have 'genuine' subject clauses in the sense that complements are treated like canonical S or A arguments. Of course, it must be added that not all two-participant events in complementation involve mental predicates and thus present analytical challenges of the type just discussed. Two-place utterance verbs corresponding to 'say', for instance, evidently involve a human agent as the initiator of the utterance and the contents of the statement as a patientive entity relative to that. Therefore, even though Munro (1982) and others have argued that utterance predicates are significantly different from the transitive-event prototype (e.g. in terms of the 'affectedness' of the undergoer), there is much less of an analytical headache in relation to determining 'A' and 'P' in quotative constructions. Similar remarks apply to causative, jussive, phasal and other verb classes.

The discussion so far has focused on one- and two-argument clauses, but Croft's model from above also includes three-argument clauses as syntactic frames for three-participant events. Prototypical events of this kind involve the physical transfer of a possessed item or the mental transfer of an experience. In complementation, the corresponding clause patterns become particularly relevant in the verbal transfer of information, i.e. when the complement clause codes the content of an utterance that is directed at a recipient, either in a declarative way ((16a) below) or a jussive way ((16b)):

(16)   a. *My cousin told me* [*that he had given up smoking*].
       b. *My cousin told me* [*not to worry about his illness*].

The transfer event can also involve physical force, which becomes relevant for causative complements in three-argument clauses:

(17)   *My cousin forced me* [*to write a letter*].

As can be seen in Fig. 1 above, it has been customary in typology to use the label 'T' ('theme') for the transferred argument, and 'G' ('goal', or endpoint in Croft's model) for the recipient, addressee or experiencer. These are again, therefore, labels for generalized semantic roles because they apply regardless of whether G is

morphosyntactically aligned with the P argument of a transitive clause (as in (16) above) or with oblique arguments (cf. Dryer 1986 for a seminal study of the corresponding alignment systems). Moreover, as Haspelmath (2011a: 559) points out, it is considerably more difficult to define a canonical type of three-argument coding across languages than for two-argument clauses, since even the most typical or frequent three-actant verbs like *give* allow for several coding patterns within the same language. Therefore, we will not distinguish between canonical or non-canonical coding of T arguments in the present study, recognizing simply that complement clauses can function as T in three-argument clauses.

Now that the basic argument-structural frames are established, the final set of concepts to be introduced into the discussion relates to the terms 'subject' and 'object' in cross-linguistic research, which I have already used occasionally in this chapter but not yet characterized properly. Following Andrews (2007a), I conceive of S, A, P, T and G as **syntactic functions**, while 'subject' and 'object' are labels for **grammatical relations**, i.e. particular alignment patterns of the syntactic functions. Thus the traditional notion of 'subject' captures the alignment of (canonical) S and A into a syntactic category for specific purposes, as opposed to the P and T or G arguments, which are referred to as 'object'. It has often been maintained that this kind of terminology cannot readily be applied to ergative-absolutive processes of alignment, which cluster S and P into a category but treat A differently. This is perhaps the single most important reason why it has become customary to operate with the labels S, A, P etc. in the first place. In accordance with this practice, the analyses in the present study will thus also be conducted and discussed for the syntactic functions {S}, {A} and {P/T} separately, bearing in mind that these may not always be meaningful syntactic categories for individual languages (in contrast, for example, to the grammatical relations established by the clustering of {SA} or {SP}). However, it will sometimes be convenient to also have an overarching comparative label for complement clauses that are associated with the traditional categories subject and object, respectively. We will see, for instance, that it can be fruitful to compare the structure of complements in {S} and {A} function as opposed to that of complements in {P/T} function. For these purposes, I shall call the former '**subject clauses**' and the latter '**object clauses**', regardless of whether these are meaningful descriptive categories in individual languages.

2.2.2 *On the notion of 'argument'*

The present section rounds off our discussion of the argument-structural dimension of complementation by considering the concepts of 'argument' and 'valence', which have hitherto been taken for granted. In the onomasiological approaches mentioned at the very beginning of this chapter, researchers speak of complementation when a predication acts as a **semantic argument** of another predicate. Conceived of this way, an argument completes the sense of a predicate or, in cognitive-linguistic terms, elaborates the substructure of a predicate (Langacker 1987: 304). Arguments contrast with **adjuncts** in that the latter provide circumstantial, peripheral information that is neither affected by nor itself affects the sense of the predicate. While this distinction

between semantic arguments and adjuncts appears to be straightforwardly dichotomous at first sight, it has long been noted that the underlying concept of **semantic valence** is a gradient one. For example, it is intuitively clear that a predicate like *kill* has two "salient substructures" in Langacker's (1987: 300) terms, i.e. a killing entity and a killed entity. We can thus be confident that *kill* has two semantic arguments. However, it is less clear whether a predicate like (*be*) *angry* involves two substructures that are both as salient as the ones of *kill*; put differently, does *be angry* entail reference to just an experiencer of the emotion or also to the source of the anger, and does it do so to the same degree? The fact that, in English and quite a few other languages, both an adverbial clause and a complement clause are possible expressions for this substructure (*I am angry* [*that/because she left*]) may be taken as an indication that this is not the case, and whence as a reflex of the gradient rather than absolute nature of the semantic argument-adjunct distinction. Semantic approaches to complementation thus face the problem of deciding where the conceptual domain of complementation ends and that of adjunction begins.[10]

On Dixon's approach, by contrast, the distinction between arguments and adjuncts takes place on a *syntactic* level: In order for a grammatical construction to qualify as a genuine complement clause, it must be able to function as a **syntactic argument** of a predicate. Being a semantic argument (by whatever criteria) is not sufficient; it is the morphosyntactic treatment that counts. Across many languages, a recurrent syntactic property of arguments is their obligatory occurrence, or non-omissibility, which contrasts with the relatively free addition (and omission) of syntactic adjuncts (also known as adverbials). However, it is generally agreed that obligatoriness is "a sufficient, but not a necessary condition for argument status" (Comrie 1993: 907): Phrases and clauses whose omission leads to ungrammatical sentences can be considered arguments, but the converse does not hold: the optionality of an element in a sentence does not automatically rule out argument status. As is well known, many languages freely omit core NPs from clauses if they are sufficiently retrievable from the discourse context. In such cases, a grammatical distinction between arguments and adjuncts may be maintained by other morphosyntactic criteria. In head-marking languages like Tzutujil, for example, arguments are indexed on the verb and the coreferential NPs can be freely omitted if their referents have been given in the previous discourse. Thus (18) below would be a grammatical sentence without the terms in brackets:

(18)   Tzutujil (Mayan: Guatemala; Dayley 1985: 299)
    Xkeetij   *(ntzyaq)*   *(ch'ooyaa?).*
    3ABS.3PL:ERG.ate   (my.clothes)   (rats)
    '(Rats) ate (my clothes).'

---

[10] However, as Deutscher (2000: 9) points out correctly, this apparent problem also has an important advantage: Accepting the gradient nature of semantic valence, i.e. the seamless transition from adjuncts to arguments, opens the door to understanding the historical dynamics of complex-sentence systems. It is precisely the grey area between more or less salient substructures that provides the historical bridging contexts in which, for instance, an adverbial clause ('angry [because she left]') is reanalysed as a complement clause ('angry [because she left]' > 'angry [that she left]'). We shall return to these processes in more detail in Chapter 7.

In the literature, it has sometimes been suggested that the cross-referenced NPs in head-marking languages should be considered adjuncts because of their omissible status (most famously, perhaps, by Jelinek 1984). However, as many typologists (e.g. Siewierska 2001, Croft 2001) have argued, these NPs have very little in common with what is normally considered an adjunct; moreover, "in many languages with cross-[referencing] the conominals are indistinguishable from nominal arguments of languages without cross-[referencing], and it seems ill-advised to say that they are not arguments." (Haspelmath 2013: 209) In languages like Tzutujil, therefore, the syntactic argument status of NPs is not defined by their obligatory presence in the clause, but by their being treated in a particular way by indexation: The normal pattern in head-marking languages more generally is that only (certain subsets of) arguments, but not adjuncts, are indexed. By extension, if a complement clause is cross-referenced on the matrix predicate in the same way that an object NP would be, it functions as an argument of the CTP, and thus needs to be considered a genuine complement clause in Dixon's sense. This arguably holds for complement clauses in Tzutujil:

(19) Tzutujil (Mayan: Guatemala; Dayley 1985: 399)
      Xinwaajoʔ   [*chi   xwari*].
      3ABS.1ERG.wanted   COMP   3ABS.slept
      'I wanted him to sleep.' (lit. 'I wanted it that he sleep.')

Apart from (or in addition to) obligatoriness and indexation, there may be other criteria to separate arguments from adjuncts syntactically. Especially in isolating languages without any case marking or indexation and with the possibility of argument omission, more subtle tests are sometimes necessary to distinguish between argument and adjunct clauses.[11] In such cases, I generally went with the authors' judgments on argument status, and classified the complement accordingly.

However, sometimes the picture is complicated by the fact that, in the absence of any clear morphosyntactic criteria for argumenthood, obligatoriness is virtually the only clue that one could turn to, and opinions on whether a given element is obligatory or not may be divided between analysts of the same language. The decision is often heavily dependent on the methodology being used and the register being considered, and it may be influenced rather strongly by prescriptive norms in a linguistic community. This is nicely demonstrated by Englebretson's (2003) monographic study of complementation in Indonesian, which I would like to discuss in the remainder of this section because it illustrates some rather deep-rooted problems that cross-linguistic research on complementation has to face.

Many accounts of standard Indonesian (which has a strong prescriptive tradition) recognize a syntactic function called 'complement', which comprises non-subject constituents that are normally obligatory in the clause but do not evince the

---

[11] To provide two examples from the literature: In Goemai (Afro-Asiatic, Chadic: Nigeria), arguments (and genuine complement clauses) precede certain types of markers in the clause, such as the progressive aspect morpheme *yi*, while adjuncts follow them (Hellwig 2006: 210). In White Hmong (Hmong-Mien: China, Vietnam), only arguments (and genuine complement clauses) can be targeted by certain topicalization processes (Jarkey 2006: 120). In both languages, other (more explicit coding or behavioural) criteria for argument status are not available or not applicable.

morphosyntactic behaviour of canonical objects (e.g. no passivization, applicativization, etc., cf. Sneddon (1996), who considers such structures intransitive verbs with an obligatory complement). Crucially, many of the predicates in question are precisely the ones associated cross-linguistically with complementation, such as 'know', 'say', 'think', etc. When these predicates co-occur with a clause, this clause may thus be said to fill one of their argument positions. However, based on extensive data from spoken colloquial Indonesian discourse, Englebretson takes issue with the alleged obligatoriness of the 'complements' of such verbs. Oftentimes, either the putative matrix clause or the 'complement' can be found on its own or in a separate intonation unit (e.g. as a prosodic 'increment' in Ford et al.'s (2002) terms), and there are no reliable morphosyntactic criteria that distinguish such alleged complements from adjuncts. The predicate *tahu* 'know', for example, may form a grammatically and prosodically complete clause without a complement, and cases that resemble complements in other languages may still not have argument status in Indonesian:

(20)   Indonesian (Austronesian, Malayic: Indonesia; Englebretson 2003: 52)

    IU1:   *Nggak* *soal-nya*   *udah*   *tahu,*
            NEG   problem-*nya*   already   know

    IU2:   *dia punya duit   banyak*
            3SG own   money much

    'No, the thing is, they knew she had a lot of money.' (IU = intonation unit)

Overall, the conversational data appear to suggest that Indonesian speakers juxtapose unmarked clauses to predicates that are more or less self-sufficient 'framing elements' for the content of the juxtaposed clause; they may be 'projective' in a discourse-pragmatic sense (e.g. Ford and Thompson 1996, Auer 2009), but do not open up genuine argument slots that one could analyse as grammatical complements. Their seemingly obligatory status in Indonesian grammar

> "may be true for decontextualized sentences" in the standard variety, "but the notion of semantic incompleteness is not defensible for language viewed in its discourse and social context. […] What makes an utterance complete or incomplete is its relevance to the social and discourse context. Thus to characterize complement clauses as something without which a sentence would be 'incomplete' is neither a useful definition nor an accurate reflection of the data" in Indonesian. (Englebretson 2003: 64–65)

Consequently, Englebretson denies the complement status for the vast majority of the clauses in question, at least for the variety under scrutiny. Interestingly, the same point is made, based on similar evidence, by Thompson (2002) for unmarked finite clauses in conversational English.

    The problem that arises from such studies is thus: Which genre or variety should be taken as the basis for assessing the argument status of the complement? Is 'non-complementation' in the colloquial spoken variety a suspension of a standard grammatical pattern, or does it rather need to be recognized as the *basic* pattern from which the codified standard variety has arisen as a "normativized specialization" (Thompson 2002: 142)? This is a non-trivial question, especially if we subscribe to the

usage-based view that grammatical structure emerges directly from language use. However, the present study is more conservative in that I adopted the analyses provided in the materials at hand, and these materials tend to be framed in more traditional grammatical descriptions than so-called 'interactionist' theories of grammar (such as Schegloff 2001, Thompson and Couper-Kuhlen 2005). For example, if Englebretson's study had not been available, my policy would have resulted in adopting Sneddon's analysis, according to which (at least standard) Indonesian does have complement clauses. In other words, the coding of the data in the present study may sometimes be based on a more traditional understanding of 'obligatoriness' and 'complementation', and run the risk of distorting the full empirical reality in a given language.

And note a further problem: Even if I had based the decision on colloquial Indonesian only, I would have had conflicting evidence as to the argument status of the clauses in question. This is because Englebretson *does* find a certain amount of cases in the conversational data to which a genuine complement analysis may be applied. These concern, firstly, clauses that appear with intransitive predicates and no other argument can be found in the immediate context, and, secondly, transitive predicates that index the complement clause by voice coding or applicativization:

(21)   Indonesian (adapted from Englebretson 2003: 81)
Jadi   **di**-*kira,*   *… apa?*   *Udah*   *habis.*
so   UG-guess   what   already   finished
'And so it was assumed, what's-it?, (that) it's finished.'

As for indexation, Englebretson claims that the verbs affected (e.g. 'say' or 'guess') are atypical transitive verbs, both cross-linguistically and in Indonesian (where they never occur with an object NP), and so their rare transitive coding with a juxtaposed clause (as in (21) above) must be due to something other than complementation being at work. For example, 'say' with an index for undergoer voice is claimed to be "an evidential marker with a fixed form and no argument structure at all" (Englebretson 2003: 81). Similarly, the clauses co-occurring with applicativized verbs are typically ditransitive and are indeterminate as to whether the recipient or the theme is promoted to object status. Both possibilities are grammatically sanctioned, as careful elicitation shows (ibid.: 77ff., 83ff.), and so it remains open whether it is the complement clause that is indexed. However, in a few cases the juxtaposed clause is the only non-subject entity in the sentence and so the applicative suffix must be targeting it; therefore, "at least 7 of the framing verbs suffixed with [the applicative] *-kan* take a clause as complement" on grammatical grounds (ibid.: 87). Overall, "based on verbal morphology, 11 of the 263 examples […] in the database, a mere 4.18%, appear to be complements – a clause functioning as a grammatical argument." (ibid.: 88) The critical question now is whether these minority cases are enough to justify that juxtaposed clauses should be elevated from complementation strategies to complement clauses. Englebretson answers this question in the negative because he would consider this move "empirically irresponsible" (ibid.): If 89-96% of all instances in the data "simply consist of juxtaposed clauses related semantically, but with no grammatical connection, overt or otherwise" to the matrix predicate, "Indonesian seems not to have

grammatical complementation as a robust category or as a grammatical resource for speakers in everyday interaction." (ibid.: 89) Following Hopper's ideas on 'emergent grammar', Englebretson claims that the juxtaposed clauses which *are* treated as arguments in the corpus have not achieved a high degree of 'structuralization', "in the sense of cross-textual consistency, and serving as a basis for variation and extension" (Hopper 1987: 150).[12]

However, while I am sympathetic to the general approach taken by Englebretson (and to most of his specific analyses), I adopted a different policy in my data: Whenever there was evidence that a complementation pattern *can* function as a proper syntactic argument of the main clause (e.g. being considered obligatory for at least some predicates, or triggering the relevant indexation marking on at least some CTPs), I usually granted it complement-clause status. This practice acknowledges the fact that some syntacticization of a complementation relation has taken place, even if it is not extended (yet) across the board to all CTPs that co-occur with the construction. Overall, then, our cross-linguistic definition of complementation should take into account that a complement needs to *be able to* function as a syntactic argument even though it may not always do so in all environments in which it occurs.

This concludes our relatively detailed survey of the argument-structural properties of complementation constructions. I have maintained that complementation patterns are considered complement clauses in Dixon's sense if there is language-particular evidence (from obligatoriness, indexation, case marking or behavioural properties) that they function as a syntactic argument of a at least a subset of the predicates they co-occur with. In this capacity, they may either replace a phrasal argument (NP, PP) or function directly as arguments of predicates that do not take a corresponding NP or PP to begin with. The specific syntactic functions that may be fulfilled by complement clauses so defined have been characterised as S, A and P/T in canonical and non-canonical clause patterns, noting that my usage of these labels is broader than in classic typological research. In the next section, we turn to Dixon's second criterion for complement-clause status, which zeroes in on the notion of 'clause'.

2.3 On the notion of 'clause'

The issue of 'clausehood' actually arises from two different perspectives, each of which is fundamental to delimiting the scope of the present investigation. First, we will establish complement clauses (and complementation strategies) as fully biclausal units and exclude functionally related constructions that do not typically qualify as such. Second, we will recall Dixon's definitional criterion from above that complement clauses must exhibit the internal structure of a clause rather than that of an NP. Accordingly, 'clausehood' in this sense refers to the ways in which the syntactic

---

[12] Englebretson goes on: "In some languages, [this] overt marking [of clauses as] core arguments is much more robust, and speakers have grammaticized the co-occurrence of framed instantiations to such an extent as to mark them overtly as complements. In the colloquial Indonesian data, however, grammatical complementation is only [...] observable after the fact by analysts searching for traces of grammatical complementation, but it is simply too rare to be a viable grammatical category for language users." (89).

elements of clauses, i.e. the predicate, its arguments and potential modifiers, are expressed. Each of the two issues will be dealt with in turn.

2.3.1 Complementation constructions as biclausal units

What all the data points in the present study have in common is that they are **biclausal** syntactic entities. A clause is commonly regarded as "the smallest grammatical unit that can express a proposition" (Kroeger 2005: 32), by virtue of being organized around a predicate and its specific argument structure.[13] Therefore, each clause entering a biclausal complementation construction must be capable of establishing its own argument structure, rather than being fully subsumed under the argument structure of another predicate. The latter situation typically obtains in auxiliary-verb, serial-verb and compound-verb (or 'co-lexicalized' (Givón 2009: 62)) constructions. All of these three patterns are widely considered as monoclausal from a synchronic point of view[14], even though their historical source may well be a biclausal complementation construction (cf. Anderson 2006: 320ff. for an overview of this diachronic pattern). One important morphosyntactic aspect of this monoclausal synchronic nature pertains to the integration of the auxiliary with the argument-structural frame of the main verb. In the Amazonian language Trumai, for instance, the concept of 'desire' is not expressed by a lexical verb taking a clausal complement, but by the Auxiliary verb *take* (or its phonologically reduced form *(-)tke*), illustrated in (22) below.

(22)   Trumai (isolate: Brazil; Guirardello 1999)

    a.   *Ha-ø   xom   take   ka_in   fi-s.*
        1-ABS   suck   want.AUX   FOC/TNS tobacco-DAT
        'I want to smoke.' (132)

    b.   *Asix-ø   yi-ake   naha-tke   men   Atawaka-k.*
        tail-ABS   *yi*-3POSS   cut-want.AUX   frustratively   Atawaka-ERG
        'Atawaka wanted to cut its (a bird's) tail.' (146)

As can be seen, the argument-structural configuration of the clause, i.e. the case marking on the NPs, is entirely determined by the main verb: *xom* 'suck' in (22a) takes an Absolutive-S argument and a Dative-P argument; *naha* 'cut' in (22b) is a canonical transitive verb showing an Ergative-A and an Absolutive-P argument, respectively. The Auxiliary 'want' has no influence on the case marking patterns. In genuine complementation constructions, by contrast, the matrix predicate is a regular main verb and hence capable of establishing its own "independent argument structure" (Guirardello 1999: 136). Furthermore, while the Particle *yi* can be inserted between the predicate of a complement clause and the matrix verb, no such separation of

---

[13] cf. also Bossong (1979: 34): "Die den Satz als kleinstmögliche Äußerung konstituierende Beziehung nenne ich *Fundamentalrelation*. Sie ist gekennzeichnet durch die spezifische Verbindung eines Valenzträgers, des Prädikats, mit dem oder den ihm entsprechenden Aktanten."

[14] In a recent survey of auxiliaries, for example, Anderson (2006: 5) considers an auxiliary "an element that in combination with a lexical verb forms a monoclausal verb phrase". Similarly, a typical serial verb consists of "two or more verb roots which are neither compounded nor members of separate clauses" and often follow an intonation contour that is "characteristic of a single clause" (Payne 1997: 307–308).

desiderative *take* and the main verb is possible. These two properties of complementation that distinguish it from auxiliation are illustrated in (23) below:

(23) Trumai (isolate: Brazil; Guirardello 1999: 137)
    Hai-ts   *chï̱_in*   [*axos waṭkan yi*] *homne.*
    1-ERG   FOC/TNS child cry   *yi*   find
    'I found the child crying.'

The predicate *homne* has an Ergative A-argument and takes an Absolutive (zero-marked) clause as its complement; if it were an auxiliary, the first-person subject pronoun would appear in Absolutive case (triggered by intransitive *waṭkan*), and it would form an inseparable unit with *waṭkan*. Auxiliary-verb constructions in Trumai are thus clearly different from complementation constructions and can hence safely be excluded from the present study.

    With some of the sample languages, however, the difference is less clear-cut, often because the auxiliary has directly grammaticalized from a complement-taking predicate and thus forms a synchronic continuum with it. As Bolinger (1980: 297) already noted, "the moment a verb is given an infinitive complement, that verb starts down the road of auxiliariness." For the Mayan language Tzutujil, for example, Dayley (1985: 404) remarks that verbs taking Infinitival complements "function much like auxiliary verbs" in other languages, but they behave syntactically like regular complement-taking verbs; therefore, the Infinitive construction can fully be taken into account as a complement clause for the predicates in question. For Dolakha Newar, Genetti (2006: 151–53, 2007: 421–22) shows that an Infinitive construction functions as a genuine complement clause of the impersonal predicates 'be necessary' and 'be appropriate', i.e. as S-clauses, but that an auxiliary analysis has to be adopted for predicates that, at first glance, seem to take the Infinitive as their object clause. Thus phasal, conative, abilitative ('can, be able') and some other predicates appear in structures like the following:

(24) Dolakha Newar (Sino-Tibetan, Tibeto-Burman, Bodic: Nepal; Genetti 2006: 153)
    Āmun   *ām=ri*   *sugā=ta*   *syār-i*   *mwāl-ai.*
    3SG.ERG that=IND parrot=DAT kill-INF   try-3SG.PRS
    'He tries to kill the parrot.'

As can be seen, the sentence-final predicate *mwāl* 'try' indexes the subject of the matrix, just like a regular CTP. Other than that, however, such predicates "do not have any effect on the morphosyntactic structure of the clause" (Genetti 2006: 152). In particular, Genetti claims that the case marking of the subject NP in the clause (here: *āmun*) is determined entirely by the lexical verb (here: *syār*) rather than the putative CTP[15]; consequently, the resulting structure should be seen as a simple sentence involving a main verb and an auxiliary. On this analysis, there is only one object in

---

[15] Unambiguous examples are hard to come by since hardly any of Genetti's example sentences involve an overt clash in case requirements between the two predicates involved (e.g. the lexical verb requiring an Absolutive subject but the auxiliary requiring an Ergative subject, or vice versa), although such contexts do exist according to the description (*phar-* 'able' may be a case in point (cf. Genetti 2006: 152, ex. (28)), but the transitivity of this auxiliary is not discussed explicitly).

(24), viz. the phrase *ām=ri sugā=ta*, rather than a clausal object [*ām=ri sugā=ta syār-i*]. This is corroborated by evidence from linear order, similarly to Trumai above: In simple sentences, it is normally possible to swap the order of subject and object, yielding OSV as an alternative ordering to the regular SOV. This option is also available for complement clauses but, crucially, not for infinitival constructions with the above predicates. Thus when *lumonker* 'forget' functions as a lexical verb in a simple sentence, both (25a) and (25b) are possible ordering patterns; but when combined with an Infinitive, the ordering in (25c) is ruled out. Therefore, Genetti suggests, "the primary lexical verb" *ŋyār* "plus the auxiliary" *lumonker* "form a single verbal complex" (Genetti 2007: 421):

> (25) Dolakha Newar (Sino-Tibetan, Tibeto-Burman, Bodic: Nepal; Genetti 2006: 153)
>
> a. *Jin*   [*ām khã*]   *lumonker-gi.*
> 1SG.ERG   that talk   forget-1SG.PST
> 'I forgot that talk.'
>
> b. [*Ām khã*]   *jin*   *lumonker-gi.*
> that talk   1SG.ERG   forget-1SG.PST
> 'I forgot that talk.'
>
> c. *[*Khẽja ŋyār-i*]   *jin*   *lumonker-gi.*
> egg   buy-INF   1SG.ERG   forget-1SG.PST
> 'I forgot to buy eggs.'

In sum, there seems to be some evidence that an auxiliary-verb analysis has to be adopted for these specific occurrences of the Infinitive. This means, however, that a given complementation pattern (here: the Infinitive) is simultaneously available as a genuine complement clause (in S-contexts) and as part of an auxiliary-verb construction (in P-contexts). In such cases, the policy adopted in the present study was such that the complementation pattern did enter the analysis as a regular data point, but that the predicate classes manifesting auxiliation were coded to exhibit a distributional restriction. (The coding scheme for the distributional potential of complements across CTP classes will be presented in §6.2.)

The phenomenon of **serialization** also requires some brief remarks. To the extent that typical instances of serial-verb constructions form a single complex predicate and hence open up one rather than two separate argument-structural frames, the same comments apply as to auxiliary constructions above. For example, the complementation system of Tariana employs a variety of genuine complementation patterns, but also features a special structure for phasal, same-subject desiderative, jussive and causative verbs: It consists of the serialization of two verbs "with no syntactic linkage, which form one predicate" (Aikhenvald 2006: 197) and are indexed for the same subject even when the subject participants of the two verbal situations are different:

(26) Tariana (Arawakan: Brazil; Aikhenvald 2006: 197)
[*Nu-ira-de*   *nu-nu*]   *di-na.*
1SG-order-FUT.CERT   1SG-come   3SG.NF-OBJ
'I will order him to come.' (lit. 'I-order I-come him.' = 'I order-come him.')

The example nicely shows how the two events are integrated into a single argument-structural frame and hence result in a monoclausal structure. While this and even more tightly knit verb complexes are perhaps the prime instances of what is referred to as serialization in the literature, the term has also been used more broadly to encompass a variety of structures in which verbs and verb phrases of different sizes are juxtaposed without any formal marking. This is particularly characteristic of languages with a low degree of morphological synthesis, such as the languages of South East Asia. In the comparative analysis of five of these languages, Bisang (1992) thus distinguishes precisely between serialization in a wider and a narrower sense. Serial constructions in the narrow sense have undergone some degree of co-lexicalization or grammaticalization (and match the classic definition of serial verbs in typology), while serial constructions in the wider sense show a much lower degree of cohesion and are subject to few syntactic restrictions (cf. Bisang 1992: 2, 36ff.). The phenomena covered by this latter sense of serialization comprise, among other things, the simple coordination of verb phrases (as in (27a) below), the modification of a verb by an adjacent one (as in (27b)) and, crucially, what Bisang (1992: 40) calls "governing serialization" (as in (27c)). In the latter construction, a VP or clause functions as the syntactic object of a preceding verb (ibid.) and thus constitutes a bona fide instantiation of complementation.

(27) Examples of serialization in a wide sense (Bisang 1992)

a. Juxtaposition in Hmong
   Nws thiaj li sau   nra   khiav.
   he   then   collect belongings escape
   'He collected his belongings and fled.' (37)

b. Modifying serialization in Vietnamese (*lón* = a verb)
   Bảo cười   lón.
   Bao laugh loud.V
   'Boa is laughing loudly.' (38)

c. Governing serialization (= complementation) in Mandarin Chinese
   Wǒ yào   tā qù shìchǎng.
   I   want he go market
   'I want him to go to the market.' (40)

Needless to say, examples such (27c) are quite different from the traditional understanding of serial verbs. In fact, in other sources on Chinese, such as Yue (2003: 115) or Lehmann (2002: 55), the construction above is simply treated as "an independent sentence that is embedded as the subject or object of another sentence without being nominalized into an NP through marking." A similar situation also holds for Vietnamese (which, like Mandarin Chinese, is one of our sample languages). For this language, too, Bisang (1992: 318) argues that many verbs can 'trail behind' a

VP or a full clause without any marking (although for some of them the insertion of a marker is possible); as here, too, this VP or clause is the object of the preceding verb, the whole construction is a prime instance of complementation. More generally, then, only serial verbs of the closely knit type are excluded from my analysis.

In sum, the present study only takes structures into account that, at least in some of the relevant environments, are to be considered truly biclausal syntactic entities. Grammaticalized, serialized or co-lexicalized variants of erstwhile full-fledged complementation (or other multiclausal) patterns are not considered as separate data points. They arguably do not constitute complement clauses on any of the conventional definitions of the term and, in contrast to Dixon (2006a: 34), I will also refrain from taking them on board as complementation strategies. While it is undeniably the case that auxiliaries, serial verbs and similar structures can be functional equivalents of complement clauses in certain semantic domains, they are typically *monoclausal* alternatives that, at least for the purposes of the present study, are considered to be outside the realm of complementation systems. Rather, they are often important diachronic offshoots of complementation constructions, whose specific developmental pathways and synchronic varieties have already been studied thoroughly in the typological literature.[16] In a sense, then, the above phenomena constitute developments *away from* complementation[17] and in particular also from some of the major questions relating to complementation systems in the present study (e.g. the specific syntactic functions that complement clauses fulfil across languages, which is a nonsensical question for monoclausal structures). However, despite the fact that auxiliary and serial constructions did not enter the analysis in their own right, we will, of course, be cognizant of their existence as alternative expressions to complementation and thence as an important factor that may help explain the restricted productivity and specific functional distribution of complement clauses in certain languages. This issue will be taken up in §6.3.

2.3.2 The internal structure of clauses

Once we have settled on the cases that can be considered predicate-argument constructions to begin with, the next issue that arises is for their internal structure. As will be laid out in detail in Chapter 4, complementation amounts to construing an action or process as a referential expression, i.e. in the typical function of an NP (Lehmann 2002: 54, Croft 2001: 88). Accordingly, complementation constructions are put under a certain tension to retain the structural properties of a prototypical verbal structure, i.e. of a canonical clause, while at the same time adjusting to the nominal environment in which they are being used. As we will see, this adaptive problem is

---

[16] Apart from the works already mentioned, reference deserves to be made to Heine's (1993) and Bybee et al.'s (1994) study of the grammaticalization patterns of auxiliaries and their conceptual motivations, as well as to important typological studies of serial verbs, such as Lord (1993) and the collections edited by Lefebvre (1991) and Aikhenvald and Dixon (2005), respectively.

[17] An important exception here is the development from serial constructions into quotative markers and more general complementizers (cf. Lord 1993: Ch.7), but this appears to occur more commonly with serializations in Bisang's (1992) wider sense (juxtaposing and governing serializations) than with tightly knit serial verbs.

solved in manifold ways across the world's languages, with numerous intermediate points between fully clausal and fully nominal modes of expression. Despite this gradient nature of the clause–NP distinction, however, Dixon (1995, 2006a) imposes a categorical distinction onto it: According to his definition from above, only sufficiently clausal (i.e. verbal) entities can qualify as complement clauses proper; strongly nominalized constructions are considered complementation strategies.

The practical problem that arises here is where to draw the line between nominal and clausal structures. The decision seems relatively uncontroversial in cases where the verb of the complement clause takes nominalizing morphology, while virtually all other elements of the complement retain the same form as in independent clauses. This type of nominalization is found, for example, in Lezgian:

(28)   Lezgian (Nakh-Daghestanian: Azerbaijan, Russia; Haspelmath 1993: 361)

Pačah.di   abur.u-z   ø   dewe-ø   žaǧur-un   bujruǧ-na.
king(ERG)   they-DAT   [ø(ERG)   camel-ABS   find-NMLZ]   order-AOR
'The king ordered them to find the camel.'

(28) contains an instance of the so-called 'Masdar' complement in Lezgian, which has the external syntax of an NP, i.e. it replaces phrasal objects and appears in their position in the main clause. However, "in its internal syntax, it behaves completely like a verb. It does not take adjective or Genitive modifiers or determiners; but it does take adverbial modifiers, and the verbal arguments appear in the same form as in finite clauses" (Haspelmath 1993: 153). What is cross-linguistically more common, as we shall see in Chapter 4, is that the overt nominalization of the internal predicate is concomitant with at least some form of "actant modification" (Bossong 1979), i.e. a change in form of the subjects and/or objects, and an at least partial acquisition of other nominal properties (e.g. the occurrence of certain determiners). But it is precisely these partial changes that result in classificatory problems. Dixon (1995, 2006a) is never absolutely precise about his criteria, but the major variables he appears to rely on are the coding of the internal object and the type of modification of the complement verb (adverbial or adjectival). Accordingly, in Dixon (2006a), he contrasts the following two examples from English:

(29)   a. cc[*John's playing the national anthem competently*] *pleased Mary.*
       b. cs[*John's competent playing of the national anthem*] *pleased Mary.*

As one can see, the bracketed units differ in the formal realization of the P-argument (a canonical, unmarked NP object in (29a) but a possessive phrase in (29b)) and in the form of the modifier (*competent* versus *competently*). In both cases, however, the coding of the internal A-argument differs from an independent clause: *John* appears in the possessive form, so this kind of actant modification is granted even in complement clauses proper. In view of the cross-linguistic data on complementation[18], I consider this a reasonable approach, and I will, therefore, adopt it for the present study as well.

---

[18] Some of the languages in the sample (such as Hausa, cf. Bagari 1972) are similar to English in employing a more nominal and a more clausal form of nominalization, the former often being called a 'deverbal noun', the latter a 'verbal noun' in the language-specific literature. And a notable difference between these two patterns lies precisely in

However, in a typological study with a relatively wide scope of languages, it is not always possible to retrieve detailed information on the internal elements of nominalization patterns, especially in regard to the adverb/adjective distinction. In these cases, only the coding of the object could be taken into account; in others, an explicit statement on the appropriate analysis of the structure in question was taken as criterial. For Yimas (Lower Sepik-Ramu: Papua New Guinea), Foley (1991: 394) states that complementation is rendered exclusively by strongly nominalized constructions that are "clearly noun phrases", and that there are no complement clauses proper. If we accept this analysis (which I do), the specific nominalization construction used for complementation in Yimas would be classified as a complementation strategy.

Before we leave the present section, mention should be made of a term that is often made in connection with the internal structure of complement clauses, viz. the notion of '**subordination**'. This term does not surface in Dixon's definition of complementation, probably for the following reason: If one understands subordination in the sense of 'embedding' (e.g. Matthews 1997: 360, Quirk et al. 1985: §14.2), then subordination is already incorporated in Dixon's first criterion: As soon as a clause functions as an argument of a higher predicate, it is by definition a constituent of the clause of which that predicate is the head (i.e. the main clause). As such, it is structurally embedded in, and hence subordinated to, the main clause. However, on a different notion of the term subordination, any clause that differs in structure from an independent clause and hence cannot normally constitute a complete sentence on its own is considered subordinate (e.g. Lyons 1968: 178). This latter conception of subordination, which has also been referred to as 'dependency' (cf., e.g., Van Valin 1984 for discussion), is not part of Dixon's definition of complement clauses, indeed. This is probably because in many languages, clauses without any overt trace of dependent morphology (such as specific nominalizers on the verb or a complementizer flagging the clause) can still be embedded directly as a syntactic argument of a higher predicate. This is arguably the case in some complement sentences from English (cf. (30a) below), which occur without the subordinator *that*, but also more regularly in languages like Warembori, where there is never a subordinator to begin with:

(30) a. *I know* [*you've been to this place before*].

     b. Warembori (Lower Mamberamo: Indonesia; Donohue 1999a: 36)
       W-or-i   [*i-nan-do*].
       2SG-give-3SG   3SG-sleep-IND
       'I made her sleep/put her to sleep.'

In light of such examples, which are far from uncommon cross-linguistically, imposing a constraint that complement clauses must show traces of dependent coding seems to be an infelicitous move. Instead, the degree to which complement clauses differ in their structure is a parameter of cross-linguistic variation in complementation systems. Therefore, the only structural requirement on complement clauses proper is that they

---

the type of modification and the coding of the internal object, while the subject appears in a non-canonical form either way. It is situations like these which make Dixon's criteria well-founded.

retain sufficient characteristics to qualify as clauses to begin with (rather than being considered NPs).

2.4 The semantic content of complement clauses

Dixon's third criterion can be dealt with rather succinctly. It states that complement clauses "will always describe a proposition", which can in turn be "a fact, an activity or a potential state" (cf. (3) above). This characterization is somewhat unusual in that it uses 'proposition' as a hypernym; many other approaches to complementation reserve this term for what Dixon calls 'fact', and oppose it to 'states of affairs' (or activities, events, occurrences, potential states, and quite a few other terms, e.g. Ransom 1986, Palmer 2001). In Functional Grammar (e.g. Dik and Hengeveld 1991), propositions and states of affairs are seen as two different types of ontological entity, which in turn relate to different functional layers of the clause: States of affairs are coded by predications (i.e. by applying a predicate to appropriate arguments), while propositions are higher-level entities "pertaining to what is said or thought about states of affairs"; therefore, "propositional content can be evaluated in terms of its truth", while a state of affairs can only "be evaluated in terms of its existence: It can be said to occur, take place, or obtain; it can be located in space and time; it can be heard, seen or otherwise perceived." (Cristofaro 2003: 110) However, despite these different terminological choices, almost all researchers agree that complement clauses are capable of expressing these two general types of entity.

The real point behind Dixon's criterion is thus actually something else: He intends to separate complementation from constructions that may look deceptively similar, but designate a rather different kind of ontological entity, namely 'things' in Langacker's (1987: 189) terms. Dixon's original formulation is that a complement "cannot refer just to a place or a time" (Dixon 2006a: 19), which is why the bracketed structures in (31) below are *not* considered instances of complementation:

(31) a. *I saw* [*where John lives*].
      b. *I know* [*when John came home last night*].

According to Dixon, (31a) and (31b) are instances of headless relative clauses omitting their head nouns *place* and *time*, respectively.[19] Accordingly, he basically concurs with Noonan (2007: 53), who states that "the grammar of these clauses is best considered along with that of other relative clauses" and not in studies of complementation. Again, I consider this a reasonable conclusion and will follow it here: Structures as in (31) will not be part of the investigation.

However, a complication arises with regard to sentences like

(32) a. *I don't know* [*who will come*].
      b. *We asked you* [*what was necessary to do*].

---

[19] The specific structures in (31) have also been called 'free relatives' (cf. Andrews 2007b: 213), and are sometimes distinguished from headless relative clauses by virtue of containing an overt [+WH] element. Genuinely headless relative clauses, on such an account, would be relative clauses in which neither an overt head noun nor any other 'nucleus-forming' (Lehmann 1984) pronominal element is present.

The bracketed structures in this example are considered "interrogative complement clauses" (Dixon 2006a: 26), despite the fact that they follow the same pattern as (31a–b) above. In the literature, the structures in (32) are commonly called 'indirect questions', which are distinguished from the free relatives in (31) in that they are not referential expressions in their own right. Syntactically, they do not form "endocentric nominals" (Lehmann 1984: 327) that can appear in place of virtually any other argument. Rather, they are full-fledged clauses that need to be specifically selected for by their matrix verbs. This difference can have further grammatical consequences. In German, for instance, only free relatives, but not indirect questions, may occupy the so-called middle field (a position available for phrasal objects but not normally complement clauses):

(33)　German (Indo-European, Germanic: Germany, Austria, Switzerland; Eisenberg 2006: 326)

    a.　*Monika hat,* [*was Manfred ausgesucht hat*]*, bezahlt.*
       'Monika paid for that which (lit. what) Manfred selected.'

    b.　**Monika hat,* [*wer Manfred besuchen wollte*]*, vergessen.*
       'Monika forgot who was planning to visit Manfred.'

Indeed, there is a considerable body of research on the grammatical properties that distinguish free relatives from indirect questions (e.g. Grimshaw 1977, Lehmann 1984: 45f., 327f., and a recent overview in Fuß and Grewendorf 2011). On the other hand, the above authors urge to point out there are also conspicuous similarities between the two types of clauses. Most notably, free relatives and indirect questions share the operation of foregrounding a nominal entity, typically by interrogative pronouns. In questions, this normally goes by the name of focus formation, while in relative clauses, this has been called nucleus formation, and according to Lehmann (1984: 327), those two operations are functionally almost equivalent (though not entirely identical).

For this reason, the present study will exclude both free relatives and indirect questions from the analysis of complementation, at least as far as the selection of constructions is concerned. In other words, a complementation pattern in the present investigation is indeed considered to designate a 'process', i.e. proposition or state of affairs, and does not involve any kind of nucleus formation (in the above sense of foregrounding a nominal entity). Cross-linguistic correspondents of English *that*-clauses and *to*-Infinitives would thus be typical instances of complement clauses. The discarding of indirect questions as distinct data points does not mean, however, that we are ignorant of the important role that these constructions play in the organization of complex-sentence systems more generally. It has been pointed out that free relatives and indirect questions do, in fact, have so much in common from a functional point of view that speakers frequently substitute them for one another (Eisenberg 2006: 325). This typically happens with verbs that allow to be complemented by either a free relative or an indirect question (cf. English *know* above). Such contexts may then act as a bridge for one construction (e.g. an indirect question) to encroach upon another functional domain (e.g. that of relativization). Therefore, indirect questions and headless relatives, while not being data points themselves, will have to be discussed again in the context of the diachronic dynamics of complement clauses (cf. §7.2).

2.5 Environments of complementation

In contrast to the semantic criterion from the last section, which aimed to separate complementation from neighbouring grammatical domains, Dixon's final criterion is meant to make a domain-internal partitioning into complement clauses and complementation strategies again. This criterion relates to the complement-taking predicates (CTPs) in the matrix of a complement sentence. Let us briefly recall Dixon's wording:

> In every language that has complement clauses, they function as a core argument for verbs with meanings such as 'see', 'know', 'believe', and 'like', for 'tell' if there is an indirect speech construction, and for 'want' if this concept is realized as a lexical verb.

As it stands, this sentence would hardly be rejected by any typologist. After all, in contrast to the other criteria, it reads primarily like an empirical generalization of what we observe across the world's languages. Crucially, however, Dixon uses this generalization to deny some attested grammatical structures the status of complement clauses, demoting them to complementation strategies. In other words, the intended logic is that if a given complementation pattern does *not* combine with the predicates mentioned above, it cannot be called a complement clause but must be considered a complementation strategy. For example, the Infinitive in Dolakha Newar that we encountered in §2.3.1 above could not qualify as a complement clause since it only complements predicates requiring an S-argument ('be necessary', 'be appropriate'); recall that its putative uses as P-clauses turned out to be instances of auxiliation, which could not qualify as complement clauses either. In other words, the criterion above is to be read in such a way that genuine complement clauses "must be in O[20] (or E) function for one or more of a set of prototypical complement-taking verbs" (Dixon 2006a: 20). With regard to Dolakha Newar, this is certainly an odd requirement: There is no reason not to assume that the Infinitive is a clausal structure that acts as a proper S-argument for at least two predicates, so it clearly fulfils all other core criteria for complement clauses. (My objection is also shared by Genetti (2006: 151), who proposes precisely that the S-clause usages in Dolakha Newar qualify as proper complement clauses.)

Similarly, there are quite a few languages in which certain complementation patterns fulfil all formal criteria for being complement clauses, and even occur as O or E clauses in Dixon's sense, but with CTP classes that happen to be outside of the prototypical array. For example, the Arawakan language Tariana, described by Aikhenvald (2003, 2006), has what she calls "purposive-marked complement clauses". In comparison to other complementation patterns in the language, they have a very limited CTP distribution; in fact, the only genuine object-clause context they cover is that of the jussive predicate 'ask (someone to do something)'. According to Dixon's definition, this limited, atypical CTP distribution would demote the construction to

---

[20] O is Dixon's label for what is referred to as 'canonical P' in the present study.

being a complementation strategy, but again the verdict given by the original investigator differs, i.e. Aikhenvald considers it to be a genuine complement clause.

In fact, it even turns out that Dixon himself is not consistent in the application of his last criterion. For 'his own' language Jarawara, he recognizes a construction in which "a clause of indirect speech functions as O argument to the transitive verb *ati -na-* 'say, ask'" (Dixon 2006b: 112). However, despite the fact that this pattern has a fully clausal internal structure, that it expresses a proposition, and that it functions as a genuine argument of a transitive verb, he still "prefer[s] to regard it as a complementation strategy since it does not relate to any of the prototypical complement-clause taking verbs." (ibid.: 113) Now, not only are we left with the same arbitrary decision as in Dolakha Newar and Tariana; we also face the conundrum of why his above criterion actually includes indirect-speech contexts as part of the definition but this is still not enough to consider the Jarawara construction a complement clause!

Taken together, I cannot help but conclude that any co-occurrence requirement with a specific set of CTPs is highly arbitrary (as compared to the other criteria for complement clauses) and that little is gained by retaining it as a dividing line between complement clauses and complementation strategies. I shall therefore stay with the definitional criteria that have been established over the last sections: Any truly biclausal structure in which one clause comes to function as a syntactic argument for at least a subset of the matrix predicates that it co-occurs with will be called a complement clause for the purposes of the present study, as long as its internal structure is sufficiently clause-like by the criteria discussed above, and as long as it designates 'processes' without nucleus formation. Biclausal structures that are functionally equivalent but fail to act as syntactic argument, or whose internal argument structure is largely nominal rather than verbal in nature, will be called complementation strategies. Dixon's CTP criterion does not partake in distinguishing these two types of complementation pattern.

But CTPs do, of course, play an important role in the present study as a whole, and also in staking out the domain of investigation. The latter holds true because any typological study faces practical limitations: it would be an infeasible amount of work for one researcher to conduct an onomasiological study of complementation that includes all possible candidates for complement-taking predicates in 100 languages. Therefore, the scope of my study will, indeed, be limited by the range of CTP contexts considered, but, unlike in Dixon's approach, these will not be ranked as to their prototypicality. In the remainder of the present section, I intend to present an overview of the CTP classes that are distinguished in the present paper (§2.5.2). But beforehand, in §2.5.1, we will have to briefly clarify the ontological status of these classes, i.e. what kinds of entity they are.

2.5.1 Predicate classes as environments of complementation

Given that the decisive characteristic of complement clauses is that they are arguments of matrix predicates, it has been customary to approach complementation from the perspective of these predicates. In formal approaches, complement clauses are thus

said to be 'selected' or 'subcategorized for' by a CTP, while in functionally-minded approaches (e.g. Halliday 1994), one sometimes finds the term 'projection' to capture this intimate relationship. Furthermore, the meaning and form of complement sentences as a whole is usually taken to be crucially determined by the semantics of the CTP. Consequently, a common procedure is to define semantically coherent classes of CTPs that behave in similar ways with respect to complementation. In the typological literature, several such classifications can be found (e.g. Givón 1980, Dixon 2006a, Noonan 1985|2007), and for the present purposes, I will take Noonan's taxonomy (which is currently also the most widely used one in language description) as a starting point.

Noonan (2007: 120–145) distinguishes 13 semantic classes of CTP, operating with labels such as 'utterance predicates', 'phasal predicates', 'desiderative predicates', and so on (cf. §2.5.2 below). Although these labels suggest relatively broad semantic groupings, it is important to emphasize that what they really capture is quite specific **constructional meanings**, i.e. meanings that arise when a complementation pattern X combines with a matrix predicate Y. For example, what we label 'knowledge' predicates here captures specifically 'knowledge of a fact', which, in English, arises when the matrix predicate *know* combines with a *that*-complement clause.[21] As Cristofaro (2008) argues, there is no reason to assume that this kind of factual meaning is inherent to the meaning of either *know* or *that*-clauses; it only arises when the two are combined into a specific **complementation construction** of English. For this reason, Cristofaro proposes that syntactic patterns such as 'infinitive complement' or 'finite complement clause' do *not* have constructional status themselves but only in combination with a specific CTP.[22, 23] Turning to our CTP classes again, we can now say that what these

---

[21] Noonan was perfectly aware of this, stating that his classes "reflect the uses of CTPs in complementation" (Noonan 2007: 120). In more traditional terms, one could say that they cover particular **senses** of language-specific predicates, but as Cristofaro (2008) shows, assuming distinct verb senses, as is done is so-called 'lexical-rule approaches' (e.g. Pinker 1989), is quite a problematic endeavour, in complementation and elsewhere (for reasons that are not necessary to go into here).

[22] This now explains why, at the very beginning of Chapter 2, I reserved the term 'complementation construction' for the schema of which an entire complement sentence is an instantiation, or particular token in language use, while the clause acting as the complement was referred to as a 'complementation pattern'. In Cristofaro's approach, complementation patterns do not (typically) have constructional status themselves, so she urges us to refrain from calling them constructions.

[23] For readers who are particularly interested in Construction Grammar, let me provide the gist of Cristofaro's (2008) proposal in this footnote, which can otherwise simply be skipped. Cristofaro argues that the meaning of any given complement sentence can neither be predicted from the CTP alone nor from the meaning of the complement clause. The former is often impossible because one the same matrix predicate may yield quite different interpretations of the sentence depending on the type of complement clause it occurs with. An utterance predicate like *say* may thus acquire a declarative interpretation with one type of complement, but a jussive interpretation with another. Hence it may be a reasonable move to say that complementation patterns themselves are directly associated with a certain meaning, not unlike the alleged meaning of argument-structural patterns such as the 'ditransitive pattern' in Construction Grammar (e.g. Goldberg 1995). However, as Cristofaro shows based on corpus data from Ancient Greek, it is often difficult to come up with an overarching meaning associated with a given complementation pattern that holds constant, or is extended only predictably (e.g. by metaphorization), across *all* the different sentences in which it occurs. For example, the Infinitive construction in Ancient Greek is used to code unrealized events in many CTP contexts, so that it is tempting to declare 'unrealized action' the general meaning of the Infinitive clause itself. However, the same Infinitive clause is also found after utterance verbs, where it is used to report a statement that the speaker is not committed to, but which nevertheless refers to a realized event. As

actually record from a semantic point of view is **similar constructional meanings that can be identified recurrently in the world's languages**. This robust cross-linguistic attestation arguably holds for most of the CTP classes that Noonan distinguishes, but it must also be emphasized that for some of them, the overall information available is so unsystematic and patchy that that they had to be discarded from the analysis.[24]

On the other hand, Noonan's classification was also expanded. This is because it identifies purely semantic classes, but in the context of the present study, the specific syntactic relationship between CTP and complement clause is also important. Along this dimension, Ransom (1986: 20) distinguishes "**subject-embedding**" and "**object-embedding**" predicates, depending on the syntactic function of the complement in the main clause. In English, for example, *be necessary* is a subject-embedding predicate, while *know* is an object-embedding predicate. This two-way contrast can be mapped directly onto our comparative concepts of 'subject clauses' and 'object clauses', respectively. It then yields, for example, a subdivision of Noonan's class of 'propositional-attitude predicates' into a subject-embedding subclass (e.g. English *be certain, be likely, be true*) and an object-embedding subclass (e.g. English *think, believe*).

Taken together, we can now see that the CTP classes of the present study are two-dimensional entities comprising a semantic and a syntactic component: They are specific constructional meanings (e.g. 'direct perception of a state of affairs') that arise when a given predicate ('see') or predicate group ('see', 'hear') combines with a complementation pattern (e.g. the Bare Infinitival Clause in English) in either a subject- or an object-relation. Henceforth, each of these multidimensional entities will be called a specific **environment** of complementation.

2.5.2 Environments studied in the present work

I will now introduce the specific environments of complementation that play a role in the analyses to come. For each environment, I shall indicate how it relates to Noonan's (1985|2007) original CTP classes, and some classes also require reference to Dixon (2006a) and other works. As a typographic convention, I will use *italics* in reference to language-specific predicates and expressions (e.g. English *want*), while single 'inverted

---

Cristofaro notes, this meaning is very far from the 'unrealized event' meaning of the Infinitive in other contexts, and cannot easily be derived from these contexts by principled pathways of semantic extension (e.g. metaphor or metonymy). To account for such cases (and to circumvent other problems of previous accounts), Cristofaro proposes that complementation constructions *as a whole* (i.e. including both a matrix clause with its specific CTP and a certain type of complement clause) are associated with specific meanings. She calls this a radically non-compositional approach to the extent that it does not assume an invariant meaning of either the CTP or the CC that somehow combine into a specific composite meaning. On this account, (i) many languages have far more complementation constructions than normally assumed, because these amount to fairly lexically-specific entities; (ii) syntactic patterns such as 'infinitive complement' or 'finite complement clause' do *not* have constructional status themselves but only in combination with a specific CTP (or a coherent group of CTPs). The latter point will become relevant again when we consider the diachronic extension of complementation patterns across different CTP classes (cf. especially §7.2.2).

[24] This concerns, for instance, so-called 'achievement predicates' (*manage, try, dare, avoid, fail*) and 'pretence predicates' (*imagine, make believe, fool into thinking*, etc.), and some subclasses of the others (e.g. the 'hope'-subclass of desiderative predicates, or the 'acquisition' subclass of knowledge predicates (*discover, find, realize*)).

commas' are used for the translational equivalents of such predicates in a cross-linguistic and hence more general sense (e.g. the notion of 'want').

I. Object-embedding environments

Phasal-P

As Noonan (2007: 139) states, phasal (or aspectual) predicates profile the inception (Engl. *begin, start*), continuation (*continue*) or termination (*finish*) of an event. The specific environment is called 'phasal-P' here since the complement clause appears as an 'object clause' in our sense (e.g. *He began* [*to chop the wood*]). In order for a complementation pattern to be recognized in this function, there must be evidence that it occurs with at least one of the above shades of meaning (i.e. it does not have to be compatible with inceptive, continuative and terminative predicates at the same time).

Causative-P/T

This environment profiles the physical coercion of an affected participant into (bringing about) a resultant situation ('make', 'cause', 'force', etc.). The affected participant is coded as an object in the matrix clause or as the subject of the complement clause. Especially in the former case, the complement clause has the semantic role of 'theme' in relation to the causative predicate (and thus often assumes the syntactic function T in Croft's typological space (cf. Fig. 1 above)). In Noonan's (2007: 136) approach, causatives are included in a broader category of 'manipulative predicates', but as we shall see shortly, there are reasons for splitting this group up into at least two distinct environments.

Jussive-P/T

Noonan treats jussive predicates together with causative ones because they both denote an act of manipulation; however, with jussive predicates this act consists in the *verbal* coercion of an affected participant into (bringing about) a resultant situation ('command', 'order', 'ask'). As 'indirect commands', jussive constructions involve an illocutionary act and in this respect resemble utterance predicates (cf. below), which may lead to a different complement-selection behaviour than that of causative predicates. They also differ crucially as to whether the dependent state of affairs is realized: its realization is entailed in causative, but not in jussive complements. For these reasons, it is warranted to set up a distinct group of jussive environments.

Desiderative-P

This environment profiles the desire for a state of affairs to obtain in the future. The CTPs involved are limited here to the cross-linguistic equivalents of English *want*. This is because other desiderative notions, like 'wish' and 'hope' (cf. Noonan 2007: 132, Khanina 2008), are not covered systematically across all sources considered (cf. also Cristofaro 2003 for the same procedure due to the same predicament).[25] The class of desideratives was, however, divided into two subgroups. A pilot study and previous research (e.g. Haspelmath to appear, Khanina 2009) had indicated that same-subject and different-subject 'want'-constructions can exhibit very different selection patterns

---

[25] Where a language has lexicalized 'want', 'wish' and 'hope' separately, only 'want' will be taken into account.

as far as their complement clauses are concerned. From a grammatical point of view, then, it is more accurate to break up the semantic unity of 'want'-predicates into two usage contexts, i.e. **same-subject** and **different-subject** constellations, because these tend to co-occur with different complementation patterns. Finally, in order to qualify as a desiderative-P environment, the complement clause (coding the stimulus of the desire) must show relatively more object properties than subject properties as compared to the experiencer: Constructions that literally code 'it desires me [to go out]' fall into a different category, namely subject-embedding environments (cf. below).

Perception-P

This environment profiles the direct perception of a state of affairs by the subject referent of the main clause. The relevant CTPs are also called predicates of 'immediate perception' (Noonan 2007: 142). The sensory mode is typically visual ('see', 'watch') or auditory ('hear'). Note that propositional uses of the same predicates (e.g. *I've heard that Frank left his wife.*) are semantically very different: they typically denote the acquisition of knowledge rather the immediate perception of an ongoing situation (cf. Boye 2010 for an overview of these different uses of perception verbs). Therefore, they will not be included in this class.

Knowledge-P

The matrix predicates of this environment prototypically comprise the equivalents of English *know* and are restricted in the present paper to declarative knowledge (e.g. *I know **that** Hannah quit her job.*). Procedural-knowledge complements (e.g. *I know **how to** fix a car.*) are again very different semantically (they rather code the modal notion of 'ability') and are hence not taken into account. Note that Noonan's (2007: 129) class of knowledge predicates also includes 'acquisition of knowledge' (e.g. English *realize, discover*) as a subclass. Due to sparsity of cross-linguistic attestation, this notion will not be considered separately here; however, in the absence of a direct equivalent of *know*, the existence of acquisition-of-knowledge CTPs will be taken as sufficient evidence for a knowledge-P environment. In all cases, the stimulus must again show object-like behaviour on language-specific grounds, the experiencer being coded as the matrix subject.

Emotional-reaction-P

This environment profiles an emotional reaction of the referent of the matrix subject towards the complement proposition or state of affairs. In English, typical matrix predicates in this group include *regret/be sorry, fear/worry, be anxious/afraid, be happy/sad*, etc. The class of emotional reactions differs in its demarcation from other proposals in the literature. On the one hand, strictly speaking, emotive predicates also include the notion of desire (cf. Ransom 1986: 142), but desiderative complements are sufficiently well-described and idiosyncratic to merit being treated separately, i.e. in our two desiderative environments introduced above. On the other hand, Noonan (2007) does not have a distinct emotive CTP class; instead, he places 'regret' and 'be sorry/sad' etc. along with subject-embedding predicates like 'be important/good/easy' into a so-called 'commentative' CTP class (Noonan 2007: 127), while 'be afraid/fear' predicates "have enough peculiarities cross-linguistically to merit dealing with them as a class" on their own (ibid.: 130). In the present paper, I shall draw the boundaries

differently: First, I would like to sever typical subject-embedding from object-embedding contexts and hence I must break up Noonan's 'commentative' class and deal with 'be sorry/regret' etc. and 'be good/easy' etc. separately. Second, due to unsystematic coverage across the sample languages, it is impossible to maintain distinct emotional-P classes for the evaluative 'regret/happy/sad'-type and the apprehensive 'fear'-type. For this reason, they will be conflated into a more general emotional-reaction-P class, and either meaning is regarded as sufficient evidence for the class to be attested with a given complementation pattern.[26]

Propositional-attitude-P

The relevant CTPs profile "an attitude regarding the truth of the proposition expressed as their complement" (Noonan 2007: 124), and are limited here to 'positive' attitudes of this kind (i.e. 'think', 'believe', 'suppose', 'assume' but not 'doubt' or 'deny'). Since propositional attitudes can also be rendered impersonally (e.g. *It is likely/certain/possible* [*that Tom has left*]), i.e. as subject clauses, an important property of the current environment is the presence of an overt experiencer, coded as the subject of the main clause (e.g. *I think* [*that Tom has left*]).

Utterance-P/T

This environment profiles "a simple transfer of information initiated by an agentive subject" (Noonan 2007: 121); the relevant CTPs correspond to English *say* or *tell*, which can take a complement clause in P or T function (cf. §2.2.1 above). Terminologically, the resulting constructions have often been referred to as 'quotative constructions', sometimes in contrast to 'reportative constructions' that do not include an explicit referential source to which the quoted material is attributed (e.g. *They/People* say [*that Mark was killed*], cf. Spronck 2012 for discussion in a typological context). However, this terminology is used inconsistently in the literature (cf. Wiemer and Kampf 2013 for a recent critique). In the present paper, the term 'quotative' will be used in reference to biclausal constructions that conform to Noonan's semantic characterization above, without any further commitments as to the specific semiotic status of such constructions.[27] Noonan's formulation, "a simple transfer of information", restricts the domain to declarative complements; interrogative complements were excluded as data points in §2.4, while imperative complements were described as 'jussive' environments above. Cross-linguistically, the rendition of someone else's utterance often involves a certain amount of deictic and other morphosyntactic adjustments of the quoted material, resulting in 'indirect speech', but this is by no means a universal phenomenon (cf. also Güldemann 2008, Jäger 2007, Spronck 2012 for recent typological accounts of quotative constructions). Therefore, the question is whether indirect-speech constructions enjoy a privileged status in studies of complementation and to what extent so-called 'direct speech' should be considered at all. This issue has been discussed controversially in the literature, and

---

[26] Note, however, that niche constructions specialized exclusively in the expression of fear, such as the apprehensive or 'lest'-constructions of many Australian languages (cf. Schmidtke-Bode 2009: §3.6), will not be considered as distinct data points (i.e. they will not enter the database as separate complementation patterns since they are too restricted and idiosyncratic in their distribution).

[27] For example, we will not consider whether the term 'quote' should be reserved for what Clark and Gerrig (1990: 769) call a "*demonstration* of what a person did in saying something", as opposed to what they call a mere *description* of the content of the quote [emphasis added, KSB].

since it also plays a vital role in Dixon's approach (which the present paper is based on to a large degree), we will have to consider it carefully before we move on to the other predicate classes.

Dixon (2006a) is very forthright about this problem: He proposes that only indirect-speech constructions can be considered to be part of the realm of complementation, while direct quotation is an entirely different phenomenon (although no justification for the latter assumption is provided). Consequently, in languages that rely exclusively on direct reported discourse (henceforth DRD, following Güldemann 2008), utterance verbs "lack complement clauses, and do not enter into complementation strategies" (Dixon 2006a: 10). In principle, I am sympathetic to the view that direct quotation is often markedly different, in syntactic as well as communicative respects, from complementation as traditionally understood. In particular, it has been pointed out that the communicative roles of matrix and quoted clause are commonly reversed in DRD: as Güldemann (2008: 519–20) puts it,

> the major unit of the complex construction is normally the quote, and what is commonly called the 'matrix clause' is [a] dependent tag. […It has] little substance compared to the quote, a loose phonetic bond to the quote, a linear position not conforming to that of a verb vis-à-vis its object, little elaboration of event representation (including its absence), predication operators marking the [quote] as foregrounded, [i.e. features that] all serve in one way or another to focus on the direct quote as an alien textual entity within the discourse.

This characterization echoes many previous voices in the cross-linguistic and language-particular literature[28], and Güldemann, like Dixon, thus proposes to keep DRD distinct from clausal complementation.[29]

However, I am hesitant to discard offhand DRD constructions for at least those languages in which these are syntactic arguments (i.e. obligatory constituents) of a two- or three-place verb and may, accordingly, show certain properties of such arguments. In the South American language Hup, for example, a morphologically independent direct-quote clause is embedded in the (S)OV position and the utterance verb *n*ɔ́- 'say' "takes the quoted speech as its complement" (Epps 2008: 819); therefore, the syntactic relationship between the quote and the matrix "is best analysed as one of subordination." (ibid.: 821) Similar analyses apply to some other languages in my sample.[30] At least for such languages, then, it appears odd to say, as Dixon would do,

---

[28] Similar characterizations of (direct) reported discourse had been given, for example, by Longacre 1976, Munro 1982, Haiman 1989, De Roeck 1994, McGregor 1994 and Klamer 2002.

[29] In fact, Güldemann (2008: 243, 518) makes the more radical proposal of removing all kinds of morphosyntactically independent constructions denoting reported speech and mental states from the domain of complementation as traditionally defined. He is not alone in this position: Again, it is work in functional and especially interactional linguistics (recall §2.2.2 above) that has challenged the traditional concept of complementation and maintains instead that many alleged CTPs are prefabricated formulae that 'frame' or 'project' a main clause rather than embedding a finite subordinate clause as their object argument. Perhaps the best-known work along these lines is Thompson and associates' research on the parenthetical and/or formulaic nature of English CTPs (e.g. Thompson and Mulac 1991a,b, Thompson 2002), but like-minded analyses have also been put forward for other languages (e.g. Englebretson's (2003) study from above, Huang 2003 on Chinese or Weinert 2012 on German).

[30] As another case in point where this is explicitly argued for, we can consider Dolakha Newar again. According to Genetti (2006: 148), "syntactic and prosodic evidence strongly argues that embedded direct quotes […] are syntactic objects and hence complements of the quotative verb." (The overall situation in Dolakha Newar is different from the one in Hup and other languages, though, since there is also a separate, more overtly dependent, complementation

that utterance verbs neither have complement clauses nor enter into complementation strategies. Strictly speaking, the constructions fulfil all the criteria even for complement clauses proper: They are clausal entities that function as a syntactic argument in a biclausal construction. Therefore, a reasonable compromise would be to include such constructions in our sample but to assign them the status of a complementation strategy; in this way, we acknowledge their special properties and we can easily separate them from genuine complement clauses later on. This policy leads to the following overall picture: Quotative environments in the present study (i) primarily refer to 'indirect-speech' constructions (those that are deictically or otherwise formally dependent on a matrix); (ii) take direct-speech constructions with completely independent morphosyntax into account as complementation strategies if the language in question lacks indirect means of reporting discourse and if the quoted clause acts as a syntactic argument of the matrix predicate. By contrast, direct speech that involves the juxtaposition of fully independent sentences without any trace of embedding or syntacticization will, indeed, be discarded from the analysis (and be seen as a phenomenon outside the realm of complementation).

II. Subject-embedding environments

We can now move on to the environments that include 'subject clauses', which are particularly important given the overall agenda of the dissertation. In general, a characteristic feature of subject-embedding environments is that they provide a more **impersonal** rendering of the complement sentence (cf. Sierwierska 2008 on the term 'impersonal construction' more generally). Ransom (1986: 136) aptly refers to subject-embedding contexts as '**appraisals**' of the complement proposition, and the focus of the speaker's attention "is on the appraisal rather than the person doing the appraising." Object complementation, by contrast, describes "someone's reaction to the complement proposition", and the speaker's attention is with "the person doing the reacting, usually the subject of the verb, and that person's reaction" (ibid.). In other words, the difference between subject complements and object complements is often a matter of construing the same basic semantic content from a different perspective.

Epistemic-S and *Seem-S*
This environment is the subject-embedding analogue to propositional-attitude-P contexts: The experiencer, or holder, of the propositional attitude is normally the speaker, and the proposition whose truth is evaluated is coded as a complementation pattern in a monovalent main clause. In other words, the complement is the S-argument of stative predicates such as '(be) certain, likely, probable, possible, etc.' (which can take the form of verbs or predicative adjectives in individual languages). It is common to refer to such predicates as epistemic-judgement predicates, and so the environment will be called '**epistemic-S**' here.

A pilot study showed that there is a special case of epistemic-S environments that exhibits some idiosyncratic properties in quite a few languages (cf. also Olsen 1981) and may hence be set up as a distinct category for exploratory purposes. It consists of

---

pattern conveying indirect speech. For this reason, the direct-speech construction did not have to be taken into account to begin with.)

the complement to the equivalents of English *seem* and *appear*, and where these are lexicalized as complement-taking predicates, we will speak of '**seem-S**' clauses.[31]

Deontic-S

While epistemic-S and seem-S clauses belong to the domain of 'propositional modality' (Palmer 2001: 8), other subject clauses are concerned with the speaker's attitude towards "the performance of an act" (Ransom 1986: 32) and thus with a type of 'action modality' (ibid.) or 'event modality' (Palmer 2001: 8).[32] The specific semantic subtype of action modality to be investigated here is that of deontic modality, which is concerned with the speaker's moral judgement of an act, i.e. with whether he considers it necessary, important or allowed to do. In other words, it is about obligation and permission. Where these notions are coded as objects, the resulting construction almost invariably gravitates towards auxiliation (e.g. English *should, ought to*) or univerbation (Noonan 2007: 138), which is why they were not considered as object-embedding environments in the present study. However, deontic notions can also be rendered impersonally, again with the speaker as the (implicit) judging instance. This results in an important potential environment for subject clauses, specifically of the type 'It is necessary/mandatory [to do X]'.

It is important to bear in mind that deontic modality is concerned with potential future events (cf. Palmer 2001: 8). This is crucial because predicates like English *be important* can also yield a 'commentative' or 'factive' meaning when combined with a complement, but then the complement proposition is taken to be realized or presupposed (*It was important that he apologized to Erin.*) As Van linden and Verstraete (2011) argue, such "propositional" complements need to be sharply distinguished from the truly future-oriented, non-factual "mandative" senses, and the term 'deontic modality' should be reserved exclusively for the latter. I will follow this proposal here. Likewise, the cross-linguistic equivalents of English *be possible* are often polysemous, comprising epistemic, abilitative and permissive senses (cf. Ransom 1986: 127f.). Here, too, only the permissive sense (*It is possible (for him) to leave early*) is situated in the domain of deontic modality and will hence be taken into account.

Evaluative-S

On the one hand, this environment covers precisely the 'commentative', 'factive' or 'propositional' evaluations mentioned in the preceding section, i.e. *It was good/bad that you did not take that medicine; It was odd for Henry not to call.* On the other hand, evaluative comments can also be made in reference to the performance of an act (*Driving too fast is dangerous, It was difficult for Phil to write this letter*). Especially the latter type (with CTPs like 'easy/difficult/hard') has much less of a moral flavour than the emotional reactions involving 'good/bad/odd'; the same also holds for evaluations with *be amazing/interesting/annoying* and so on.

---

[31] From a terminological point of view, one may suggest here that 'seem' is an evidential rather than an epistemic notion. In the literature, both approaches are found. To the extent that 'seem' predicates denote "an inference from observable evidence" (Palmer 2001: 24), they belong to a group of expressions that Palmer calls 'deductive' epistemic judgements, but at the same time they indicate that the judgement is based on some sort of available evidence and hence is also evidential in nature. Ultimately, "the typological category Deductive appears in both [epistemic and evidential modal] systems" (ibid.: 8).

[32] In Noonan's (2007) approach, the relevant CTPs are collectively referred to as 'modal predicates'. He thus uses the term 'modal' in a much more restricted way than is common in the linguistic literature.

Phasal-S

In a study where subject clauses play an important role, it is useful to acknowledge a separate class of phasal predicates where these are coded impersonally. Instead of saying that *Allan began* [*to read*], one could alternatively construct *It began* [*that Allan read*], which is the usual syntactic frame for expressing phasal complementation in quite a few languages.[33] Therefore, we will consider '**phasal-S**' as a distinct environment of complementation.

Other-S

This category is a mixed bag of additional environments in which a complement clause functions as the single argument in a main clause and hence appears in S-function. The relevant CTPs are semantically heterogeneous across languages, comprising notions such as 'happen/occur', 'not (being the case)', 'follow', etc., as well as detransitivized clauses like 'it is said/heard/known/ordered that X'; in the latter case, the complement appears as a so-called 'derived-S' argument of a passive construction (cf. Van Valin and LaPolla 1997: 268 for this term). Since the coverage of these environments is fairly unsystematic across languages and language descriptions, the category '**other-S**' is meant to provide a convenient cover term for collecting all further kinds of S-clauses that languages may exhibit in addition to the environments above.

A-clauses

This environment construes a two-participant event in such a way that the complement expresses the more agentive stimulus causing a (change of) mental state in an animate experiencer.[34] This results in what we called 'experiencer-object' patterns in §2.2.1 above, e.g. in sentences like [*That Frank married Janet*] *surprised/ bothered/annoyed/pleased me*. Also in keeping with the above discussion, we include in this group those instances of predicates like 'want/desire', 'think', 'know' etc. that occur in an experiencer-object construction (lit. 'It desires me [to take a break]').

This concludes our survey of the CTP classes, or rather environments of complementation, that are considered in the present study. In the analytical parts of the dissertation, it will be investigated how each complementation pattern in the database distributes over these 17 classes, yielding specific types of form-function relationships in complementation and ultimately different types of complementation systems. In contrast to Dixon's approach, however, none of the distributional facts per se will be taken to distinguish between complement clauses and complementation strategies. This distinction, as the present chapter has shown, rests on other criteria,

---

[33] In fact, one could argue that the subject-clause rendering is, in a way, the more 'basic' variant, while the object-clause variant is derived by 'subject-to-subject raising', especially where it involves an inanimate subject (e.g. *The noise began to bother John*); cf. Perlmutter 1970 and Newmeyer 1975 for classic papers on this issue.

[34] Detailed language-specific studies have noted that A-clauses can also appear with predicates of a semantically different class, i.e. those that comment on a proposition. In relation to English, Erdmann (1987: 66) calls the relevant CTPs 'Ereigniskommentarverben', which comprise predicates like *make a difference, provide evidence, reflect something, need something* etc. (e.g. [*That Tom was made chairman*] *needs explaining/ makes no difference/ provides evidence of his popularity*.). However, these are far too scarce across the sample to be considered as a distinct group. Where they occur at all, they will be subsumed under our 'A-clauses' class as a whole.

which were laid out in detail in the previous sections. An important question that the current chapter has left open, however, is: What is the basis for distinguishing different complementation patterns in a given language, and which of those patterns were taken on board for the analysis? This issue, along with a number of other methodological questions, will be dealt with in the next chapter.

3

Data and methods

In this chapter, I am going to characterize the empirical foundation and some essential methodological procedures of the present work. In §3.1, I will describe the sample of languages to be investigated and the sources of information that have been tapped. In §3.2, I will discuss the policies by which the individual complementation patterns were selected from the sample languages. The final section §3.3 is devoted to aspects of the storage and analysis of the data; most importantly, I will introduce the reader to the database which all subsequent chapters draw on and make reference to.

3.1 Sampling and sources of information

The present study is based on data from a wide range of genetically and geographically diverse languages: Specifically, it investigates complementation patterns from a sample of 100 languages, which to my knowledge is one of the largest typological data sets to have been explored in respect to complementation so far. Access to this sample, and to the corresponding reference materials, was made possible by drawing on the infrastructure of the larger project with which this dissertation is affiliated (cf. Chapter 1). This is hereby acknowledged gratefully.

Compiling a broad and representative sample for investigating complementation, and systems of complex sentences more generally, is a challenging task. The major problem is the inevitability of what Bakker (2011: 106) calls a "bibliographical bias" of the sample towards languages (or language families) that offer detailed information on all the parameters in question and, especially for the larger project, across different construction types (i.e. adverbial, relative and complement clauses). It is this uneven distribution of comprehensive descriptions that severely compromises the creation of a tightly controlled sample in the domain of clause combining (cf. Cristofaro 2003: 92 for the same problem). In view of this difficulty, the strategy for language sampling employed here strikes a balance between following a formalized sampling algorithm on the one hand, and necessary elements of what is known as 'convenience sampling' on the other (cf. also Bakker 2011 for a practical suggestion along these lines). The sampling algorithm we initially applied was the so-called 'Diversity Value' (DV) method outlined in Rijkhoff et al. (1993) and Rjikhoff and Bakker (1998). This is a

well-known and widely used technique for creating a sample of a predetermined size from any given genetic classification of the world's languages. We began by inspecting the algorithm's suggestion for a 150-language sample based on the genetic classification provided in *Ethnologue* (cf. Gordon 2005).[1] It turned out that about 40% of the genetic units suggested for inclusion could not be taken into account because of non-existent or insufficient description of the complex-sentence system (while other parts of the grammar, such as phonology and morphology, may well have been described in detail). This figure remained relatively constant even if the sample size and the genetic classification adopted were varied (e.g. to Ruhlen 1991). Therefore, in order to obtain a sample of the envisaged size (N ≈ 100), certain manual adjustments (in particular, additions) were made to the reduced version of the 150-languages proposal. In some cases, the data points discarded could be replaced by alternative ones that met the same criteria (e.g. an undescribed isolate language from Amazonia replaced by a different isolate from the same area). In other cases, the suggestions made by different genetic classifications were compared to one another and where they diverged considerably, a reasonable compromise between them was made. For example, the *Ethnologue* classification had originally suggested including 3 Afro-Asiatic languages in a sample of 150 languages; Ruhlen's classification, by contrast, suggested including 6 genera within Afro-Asiatic in an overall sample of only 100 languages. The relative contribution of Afro-Asiatic to each sample is thus very different. In this specific case, where language documentation is exceptionally detailed, and where the stock is typologically quite diverse, we felt justified to include 5 Afro-Asiatic languages in the final sample (each from a different genus, of course). Similar decisions had to be made on a case-by-case basis, the overall criteria still being broad coverage but relative independence of data points.[2] The final sample that is used for the present study comprises a total 100 languages, which are listed by genetic affiliation in Material 1 in the Appendix. A map of the geographical distribution of these languages is provided in Fig. 2 on the next page.

Given that the sampling procedure for complex-sentence systems is complicated so enormously by the availability of comprehensive reference materials, a sample based on the most pressing criterion – relative genetic independence – is bound to include other biases that have not been controlled for. Cristofaro (2003: 92), in her own study of subordination systems, notes that it is "probably unavoidable" for some of the languages to be in areal vicinity and hence potentially subject to intense contact. This is

---

[1] I would like to thank Dik Bakker for making several DV computations available to us. The sampling and data collection process took place in 2008, i.e. at the beginning of the larger project. The dates indicated here (e.g. the version of *Ethnologue* given above) reflect the state of genealogical classification at that time. The same also holds for the availability of descriptive materials. In other words, the vast majority of the reference grammars date back to 2008 or earlier. More recently published additional material (e.g. specific articles on individual languages) was, however, taken into account.

[2] Note that it was infeasible to apply more recently advocated sampling methods that rest on so-called 'intragenealogical' procedures for choosing data points from genetically diverse stocks. Bickel (2008a) proposes sampling based on a prior statistical inspection of the dependent variable within language families, but this becomes intractable where several typological variables are investigated at the same time (as in the present case), and where the bibliographical bias described above often leaves little choice to begin with. Therefore, the sampling method employed here is the more traditional, 'intergenealogical', one.

also true of the present sample, although the precise workings of areal effects in complementation cannot always be determined on the basis of reference materials.[3] In linguistic typology, it has become customary at least since Dryer (1989a) and Nichols (1992) to consider the data from the perspective of 'macro areas', which are perhaps the largest zones of possible contact and spread effects. Many issues in the typology of complementation will thus be examined with a view to macro-areal distributions, at least from a qualitative point of view. The macro areas distinguished here are adopted from the *World Atlas of Language Structures* (*WALS*, cf. Haspelmath et al. 2005 for the first edition). They are set out in Table 2 below, with the original description from *WALS* and a list of the sample languages from each area.
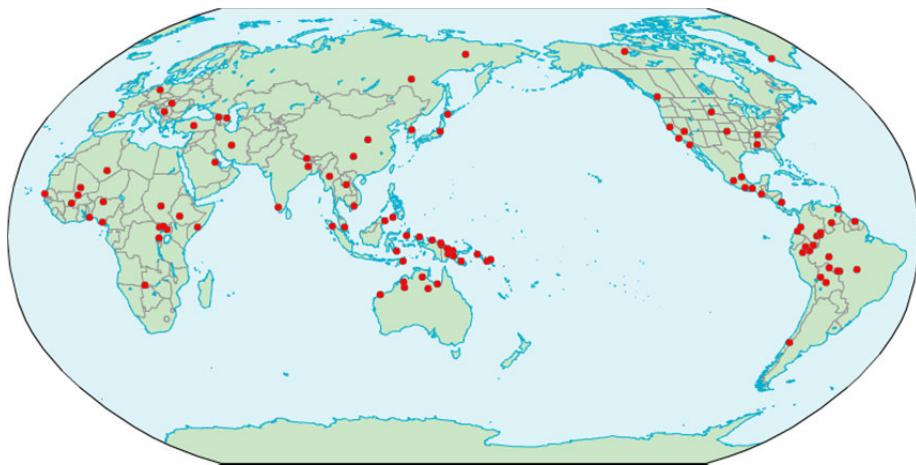
Figure 2. Distribution of the sample languages[4]

Table 2. Macro areas assumed in *WALS* and the present study

Macro area	*WALS* description	Sample languages	Sum
Africa	includes the Semitic languages of southwest Asia	Gulf Arabic, Fongbe, Hausa, Jamsay, Kana, (Modern) Khwe, Koyra Chiini, Krongo, Lango, Ma'di, Mayogo, Nkore-Kiga, Noon, Somali, Supyire, Tamashek, Wolaytta	17
Australia-New Guinea	Australia and New Guinea, excluding the Austronesian languages of New Guinea	Abun, Amele, Gooniyandi, Imonda, Kayardild, Kewa, Korafe, Lavukaleve, Mangarayi, Martuthunira, Menya, Motuna, Skou, Ungarinjin, Wambaya, Warembori, Yimas	17
Eurasia	Europe and Asia, except for South-East Asia	Ainu, Basque, Evenki, Georgian, German, Hmong Njua, Hungarian, Japanese, Korean, Lezgian, Malayalam, Persian, Serbo-Croatian, Turkish, Kolyma Yukaghir	15
South East Asia and Oceania	Sino-Tibetan, Tai, Mon-Khmer, and Austronesian	Karo Batak, Begak Ida'an, Burmese, Lao, Mandarin Chinese, Dolakha Newar, Santali, Semelai, Taba, Fehan Tetun, To'aba'ita, Tukang Besi, Vietnamese, Yakan	14
North America	includes the languages of Mexico, as well as Mayan and Aztecan languages in Central America	Choctaw, Barbareño Chumash, West Greenlandic, Kiowa, Lakota, Chalcatongo Mixtec, Musqueam, Purépecha, Slave, Tepehua, Jamul Tiipay, Tümpisa Shoshone, Wappo, Yuchi	14
South America	includes languages in Central America except Mayan and Aztecan languages	Awa Pit, Barasano, Epena Pedee, Hup, Jarawara, Kwazá, Mapudungun, Matses, Mekens, Mosetén, Ndyuka, Huallaga Quechua, Rama, Sanuma, Tariana, Trumai, Tzutujil, Urarina, Warao, Wari', Yagua, Yuracaré, Chimalapa Zoque	23

[3] For the general propensity of patterns of clause combining and clausal constituent order to diffuse areally, cf. Haig 2001.

[4] All maps in this dissertation were created with the *Interactive Reference Tool* accompanying *WALS*. Thanks go to Hans-Jörg Bibiko and the *WALS* editors for developing this tool and making it freely available, respectively.

In addition to the macro areas, certain more local areal issues were taken into account during the sampling process or will be addressed qualitatively in the individual chapters. For example, in Chapter 5 we will encounter convergence in the domain of constituent order among genetically diverse Mesoamerican languages, which is a well-known Sprachbund effect (cf. Campbell et al. 1986). Another familiar phenomenon is the borrowing of certain types of complement clauses among some of the Eurasian languages in the sample. Specifically, a finite complement clause in Turkish (*ki*-clause) and Lezgian (*xi*=clause) goes back, via borrowing, to the same Persian source. The pattern is the major type of complement in Persian, but "of very limited use in the modern standard language" in Turkish (Kornfilt 1997: 46, cf. also Bossong 1979 for the same verdict), and a direct-speech clause in Lezgian, which need not be taken into account because there is a firmly developed indirect-speech pattern as well (cf. §2.5.2 again for quotative complements). Therefore, retaining the pattern only once in the sample does not only reflect the nature of the respective complementation systems more faithfully, but also avoids a proliferation of the same structure in the data.

Overall, then, the sample for the present study aims at broad genetic and geographical coverage of the world's languages. Its basic classificatory principle is that of relative genealogical independence, but it is not a strictly controlled probability sample (cf. Bakker 2011 and Croft 2003: 21–28 for a detailed discussion of different types of samples in typology). This is because the DV method is primarily geared towards variety sampling (taking multiple units from major families into account as soon as the envisaged sample size exceeds the number of the highest-level stocks), and because of the bibliographical and areal biases outlined above. The sample is thus quite well-suited to exploring cross-linguistic variation in the grammar of complement clauses and complementation systems. At the same time, however, certain relationships between linguistic variables will be investigated in the present research. This is problematic to the extent that especially the subsamples used for exploring specific phenomena and their statistical structure may be biased even more strongly towards specific genetic units or geographical areas. However, as Cristofaro (2003: 94) notes, "one often has to make do with whatever data are available", and the problems associated with applying conventional statistical tests to a skewed dataset can be mitigated somewhat by changing the statistical strategy: Following a proposal by Janssen et al. (2006), I will employ non-parametric, distribution-free tests that rest on randomization procedures. Specifically, the statistics involve the generation of a large amount of alternative permutations of the same data and the assessment of the likelihood that the actual data are among the most skewed permutations that have been found. If this is the case, one has reason to believe that the given distribution of the data is not due to chance but represents a significantly skewed pattern. Crucially, in contrast to classic statistical methods, these randomization techniques do not make a statistical inference from the sample to the population from which the sample was drawn. They assess the statistical structure of the given data relative to possible

permutations of the data *from the same sample*[5], no more and no less. Therefore, any claim that the skewing represents a statistical 'universal' that is characteristic of the underlying population (i.e. the 'universe' of all human languages) rests on a further logical (but not statistical) inference that the sample itself is representative of that universe. As was outlined above, the present sample is intended to achieve wide coverage of the world's languages (under the given constraints) and in that sense is meant to be representative. I have to leave to the reader, however, to assess whether this kind of logical inference can be drawn from any of the statistical results reported in this dissertation.

Let us now turn to the kinds and sources of data employed for the present study. As was discussed in Chapter 2, the dissertation focuses predominantly on aspects of the grammatical structure, rather than the interactional use, of complementation patterns in the world's languages. Its primary source of data, therefore, is that of grammatical descriptions. I am fully aware of the many drawbacks that come with the reliance on reference grammars, and the limitations of this kind of data will become apparent at several points in the dissertation. Most notably, I have had to struggle with the absence of explicit information on the grammaticality or ungrammaticality of certain aspects of complementation, e.g. whether an 'extraposed' complement clause can also be found in situ, or whether a certain type of complement can be used as the subject of a transitive verb. Assessing the latter phenomenon, i.e. the co-occurrence patterns of complements and the various CTP environments established in §2.5, was a particularly challenging task, which normally required scrutinizing the entire grammar and/or text collection (as will be detailed in Chapter 6 again). Furthermore, a comprehensive understanding of complementation systems presupposes familiarization with the argument-structural properties of each language (down to the level of individual CTPs), as well with the system of complex sentences in its entirety, i.e. structural and functional links to other kinds of subordinate clauses. The latter kind of information was gathered for all sample languages (and some further ones) in the larger project on clause combining mentioned earlier, and I am grateful for being able to draw on data (e.g. on adverbial and relative clauses) collected by the project collaborators, viz. Holger Diessel and Katja Hetterle. Because of the amount and complexity of information required for research into complementation, the study of the reference grammars went well beyond scanning the passage on complement clauses. In other words, while a broad comparative study based on reference grammars must necessarily make certain abstractions of the data and cannot consider many fine-grained aspects of the structure and use of certain complement clauses, I still tried to keep an eye on language-specific and even lexically-specific details wherever possible.

In order to enhance the quality of the data, all kinds of other available material on the sample languages, or comparative literature on the stock in question, were brought to bear on the analysis. This included, inter alia, information from language-specific or comparative articles, from etymological dictionaries and from book-length overviews

---

[5] This is also called 'conditional inference', since the statistical inference "is conditional on the observed data set" (Bickel 2008b: 15). For more general discussion of sampling and statistical problems in typology, see also Perkins (2001) and Cysouw (2005).

of linguistic areas (e.g. Foley 1986 on Papuan languages, Dixon 2002 on Australia, Heine and Nurse (eds.) 2008 on Africa, etc.). Furthermore, I am particularly grateful for the support from experts of individual languages, who, due to their being either native speakers or linguistic fieldworkers on the respective language, provided invaluable comments on my enquiries. I refrained from creating and sending out a standardized questionnaire for all of the sample languages. Instead, I decided to seek the advice of experts for very specific issues that I had not found an answer to in the materials consulted, to discuss alternative syntactic analyses to the ones provided in the grammar, and to obtain acceptability judgements. Put differently, my drawing on informants was selective, and clearly secondary to the reference materials, for the purposes of this dissertation. In the future, I would like to expand this step into a more widely distributed and carefully planned questionnaire, based on the *results* and directions suggested by the present research. At this point, the study is more explorative in nature, and I wish to thank everyone who has generously responded to my query so far. The names of informants (experts and native speakers alike) are incorporated into the list of the sample languages in Material 1, and a compilation of the major printed sources that were used for the analysis can be found in Material 2. The full references of these sources can be tracked down in the bibliography at the very end of the dissertation.

3.2 Selection and nature of the data points

The next crucial methodological step after compiling a sample and possible sources of information is the choice of relevant analytical units, or data points, from the sample languages. As the title of this section suggests, there are basically two issues that need to be addressed in this connection, namely the selection process as such and the criteria which distinguish the complementation patterns selected for a given language.

The selection process was guided by the fact that the primary locus of interest of the present study is what was defined narrowly as 'complement clauses' in Chapter 2, i.e. biclausal grammatical constructions in which one clause comes to function as a syntactic argument in the other (and with the semantic restrictions discussed). Therefore, the chief targets of the selection process are the units that meet the definition of complement clauses developed in Chapter 2. However, it has also been insinuated that complementation in this narrow sense may not be a universal phenomenon, and that languages employ a wide variety of alternative grammatical techniques to achieve the same functional effects as certain complement clauses. Dixon (1995, 2006a) reserves his term 'complementation strategies' for some, though crucially not all, of these techniques. In order to systematize picture, I have created the graph depicted as Fig. 3 on the next page. As can be seen, I have borrowed the expression 'functional domain of complementation' (FDC), introduced in §2.1, from Deutscher (2000). In its widest interpretation (which may go beyond Deutscher's original intention), the FDC comprises the structures (shaded in grey here) that have already

been excluded from the analysis, i.e. monoclausal techniques and instances of direct speech that consist of juxtaposed, non-embedded discourse units.[6]
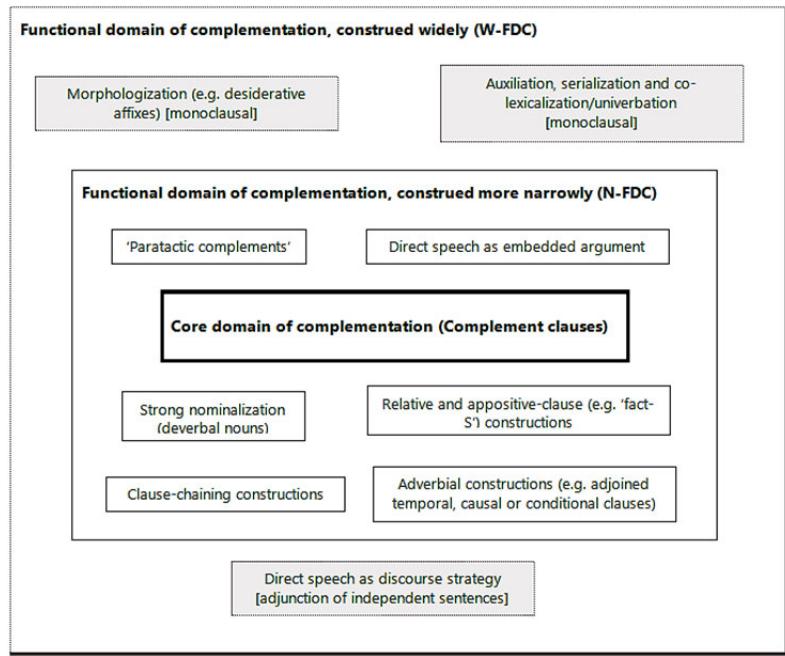
Figure 3. Delimitation of complementation techniques

In the narrower conceptualization of the FDC, we find grammatical structures that more clearly match the idea of connected biclausal constructions on the one hand (e.g. relative clauses, adverbial/chaining structures, etc.), and predications in argument roles on the other (e.g. strong nominalizations). Its core is, of course, made up by complement clauses, which combine the two properties of biclausality and argumenthood. With some exceptions (e.g. embedded direct speech), the boxes surrounding the core are Dixon's complementation strategies, and in the absence of complement clauses in the narrow sense, languages may encode a number of the CTP environments introduced in §2.5.2 by recourse to these strategies.

The gist of the selection process was to sample the **major** grammatical patterns that characterize a given complementation system, giving priority to complement clauses and adding complementation strategies from the N-FDC where appropriate. In the core domain, I tried to be relatively exhaustive. In keeping with Noonan's (2007) prediction, I also found that most complementation systems can be captured by one to three basic types of complement clause. These were considered as data points unless their description or illustration was not sufficiently detailed; some patterns thus had to be discarded because their structural properties and specific CTP profile (i.e. which environments they can occur in) remained too vague to allow for precise coding. In Chapters 4, 5 and 6, it will be seen that the complementation patterns that entered the database were analysed along a wide variety of structural and functional dimensions, and since this coding process normally requires the allocation of relatively specific and

---

[6] Note that Dixon considers serialization, but none of the other monoclausal techniques, as a complementation strategy.

unambiguous values, complementation patterns for which information on many of the variables in question was insufficient could not be considered. Also, where languages exhibit a set of basic complements and a variety of smaller patterns that are clearly restricted in their type and/or token frequency, I concentrated on the basic types, especially where the marginal patterns occur in environments that are already covered by the dominant complement types. This is what I meant by sampling 'major' complementation patterns above. For example, complementation in Evenki (Altaic, Tungusic: Russia) is characterized by the use of a 'Participial' construction type and a '(purposive-) Converbal' type, while a (borrowed) finite construction "is rather limited. It almost never occurs in folklore texts, in which the participial strategy dominates." (Nedjalkov 1997: 26) The two indigenous strategies are thus the major representatives, or the synchronic 'backbone', of the complementation system, and so only these were taken into account.[7] As we shall see in Chapter 7, other systems are characterized by a more equal division of labour between three or four important patterns, so all of them were taken into account (if well-described). In order to avoid overrepresentation of individual languages, however, I limited the maximal number of data points per language to four.

As pointed out above, complementation strategies from the N-FDC were taken on board if a language has no complement clauses on structural grounds, or if a substantial number of complement-taking environments are left uncovered by genuine complement clauses. In the preceding chapter, we mentioned Yimas as a language that does not have complement clauses in the narrow sense, but makes use of strongly nominalizing structures instead. Since these are the major complementation technique in Yimas, they must, of course, form a data point for the present study. In general, though, I was rather economical in taking all kinds of alternative strategies on board: It may be intuitively plausible that I did not want my database to be a collection including many relative or adverbial clauses. Therefore, the only complementation strategies that I took into account more regularly were strong nominalizations, such as in Yimas, direct-speech complements that are arguments of CTPs (as discussed earlier), and Dixon's appositional/adjoined complements, i.e. structures that attach to syntactically saturated matrix clauses but can still be considered complements of their CTPs. The inclusion of other complementation strategies (i.e. an occasional relative or chaining construction) thus remained exceptional and a case-by-case decision. More generally, the selection of complementation strategies was also dependent on their status in the complementation system. Strong nominalizations which have a wide distribution over our CTP classes and are described as an important member of the system (such as in Georgian) were more likely to be included than nominalization strategies which are restricted, for example, to phasal verbs. The same is true of adjoined clauses, which have a firm place as a complement type in Abun, To'aba'ita and Supyire, for example, and are not restricted to a single CTP niche. In considering

---

[7] Note that the distinction between indigenous and borrowed patterns is not generally a selection criterion. In Chimalapa Zoque (Mixe-Zoque: Mexico), for instance, the dominant and widely applicable complement clause is one with a borrowed complementizer, while other, perhaps more indigenous, strategies, are not mentioned at all in the description of the complementation system (Johnson 2000), at least not in this specific variety of Zoque.

the CTP distribution, somewhat more weight had to be given to environments involving subject clauses, as these are an important topic of the dissertation. This is why, for example, certain strongly nominalizing structures in Ndyuka, Supyire or To'aba'ita were taken into account: They are not complement clauses in the narrow sense, and they are marginal strategies in the object domain (which is firmly covered by two complement clauses in each case); but crucially, they are the only means to express certain types of subject environments, such as A-clauses (which cannot be rendered by the two complement clauses).

Overall, then, there were no hard-and-fast rules for the inclusion of complementation strategies. The choice was guided by the description of the complementation system and by the specific phenomena to be addressed in the analytical chapters of the dissertation. It must be admitted without reluctance that not a single complementation system in the sample is represented in its full complexity, i.e. with all complement clauses and complementation strategies that one may wish to include in even the N-FDC. The language-specific studies in Dixon and Aikhenvald's (2006) volume on complementation present such full-blown systematic accounts, for descriptive purposes. But given the broad comparative orientation of the present research, and the practical limits on the amount of data that can be processed within a single study, I hope that my concentration on a selection of complementation patterns is excusable. The reduction of data points is counterbalanced to some degree by the amount of information that is coded for each pattern, as will be described in §3.3 below. Also, the availability of further complementation patterns in each language has been kept track of in my records, so that it can be borne in mind when examining organizational aspects of complementation in individual languages.

Rounding off the present section, a few comments are in order on the identification of the complementation patterns as such. Reference grammars often distinguish between different 'types' of complement, which are characterized by a specific cluster of morphosyntactic properties and a particular functional profile, i.e. a distributional pattern over the CTPs of individual languages. These types are often given specific names, such as "control complement", "Infinitive", "quotative complement", etc. It appears that, across the descriptions used for the present study, the primary morphosyntactic properties for distinguishing such types are the form of the verb (i.e. whether it carries specific dependent morphology) and/or the presence of a dedicated marker of the pattern, i.e. a subordinator or 'complementizer'. These are also the chief properties by which I identified and distinguished the construction types for my database. We will see in Chapter 7 that the form of verb and the specific marker of a pattern are, indeed, responsible for one of the major areas of interest in the current study, i.e. the synchronic distribution of a given pattern (especially to different syntactic functions); therefore, it makes sense to rank these two variables highest in the differentiation of complementation types. Another important consideration, however, is the degree to which a complement type shows signs of deviations from a canonical independent clause (e.g. severe reductions in TAM or agreement potential, equi-deleted subjects, etc.). This is the perspective from which complements will be investigated in detail in Chapter 4, and so if a language distinguishes two types of

complementation patterns based on such criteria (while the verb form and the subordinator do not vary), they were also coded as distinct data points in my sample.[8] For instance, Fehan Tetun (Austronesian: East Timor) has two types of complement without a subordinator and without any dependent morphology on the verb, but they differ strongly in regard to morphosyntactic desentialization processes. Consequently, one of them is called 'sentential complement clause', the other one 'reduced complement clause' (van Klinken 1999: 278, 288). Where descriptions employed further or other criteria for distinguishing constructions, these were re-examined from the perspective of the verb form, the constructional marker, and desentialization. For example, the description of Kana (Niger-Congo: Nigeria) distinguishes three different types of complement clauses (Ikoro 1996: 280), but all of them involve the same balanced type of verb form and the same subordinator, and none of the types is markedly desentialized as opposed to the others. The structural differences between them are of such a minor nature when compared to other distinctions in complementation systems (and can easily be captured by the variables to be introduced in Chapter 4) that we are justified to speak of a single type of complement, especially for comparative purposes. Conflations of subpatterns were also applied in other cases; the major recurrent situations are listed in the following:

- A certain pattern exhibits more than one possible subordinator, but the basic type of construction, including the behaviour on all structural variables, is identical. This happened recurrently for nominalized constructions. In Matses, for example, up to 12 different Action Nominalizers can occur in complementation, but each of them is applicable to only a very limited set of CTPs, and the technique of nominalizing clauses contrasts *collectively* with other grammatical mechanisms used to create complements (e.g. an Infinitival construction and a specific Converb, cf. Fleck 2006). For this reason, I treated the nominalizations in Matses as a single construction type or 'macro-pattern' of complementation. The same lumping strategy was adopted for other languages with similar problems (e.g. Evenki, Malayalam, Mapudungun, Tariana, Kolyma Yukaghir, etc.).

- A related though slightly different case is presented by languages in which it may well have been advisable to set up several distinct patterns but where the information on their exact properties and CTP distributions was not sufficient for doing so. For example, the constructional profiles of the Krongo Infinitive and Nominalization patterns are difficult to tease apart. As Reh (1985: 333ff.) demonstrates, the two patterns are formally identical in some environments. For this reason, and because the specific distributional characteristics of the Infinitive remain somewhat opaque, I conflated the two variants into a single macro-pattern of complementation (which, as a whole, contrasts with a more balanced complementation pattern used in indirect speech).[9]

---

[8] Note, however, that if the difference between them boils down to variation between two types of overt TAM expression, such as Indicative versus Subjunctive mood in Lango (while the rest of the pattern is structurally identical and neither variant is markedly more 'desentialized' than the other), only one basic complement type was identified. (In other cases, the distinction between Indicative and Subjunctive complements also entails a different subordinator (e.g. Fongbe) or clearer signs of morphosyntactic reduction (e.g. Supyire), so that I did, of course, take both the Indicative and the Subjunctive on board.)

[9] Incidentally, it was not uncommon that a lexical nominalization ('verbal noun') is known to exist in a given language, but that its specific structural properties and particularly its contribution to the complementation system (CTP coverage) could not be retrieved satisfactorily from the sources, so that it could not be taken into

- Another complication arises when a certain construction is described as a unified grammatical pattern with a specific complementizer, but a limited number of CTPs cannot occur with the subordinator in question (though the structural pattern itself remains the same). In general, I suspect that the systematic absence of a complementizer with certain CTPs points us towards a *distinct* complementation pattern. In Tamashek (Berber: Mali), for instance, a completely unmarked finite clause acts as the complement of a host of different predicates, and there is no evidence that this pattern is synchronically or diachronically related to the complements with an overt subordinator (Heath 2005: 687). However, there are also cases in the sample in which the absence of the complementizer is synchronically limited to a very small number of CTPs, so that it becomes questionable whether these should be credited with a separate complementation pattern. Consider the Atlantic language Noon (Senegal). In this language, there is a "finite complement clause […] normally linked to the main clause by the complementizer *an*" (Soukka 2000: 274). Crucially, different-subject desiderative environments are special because they cannot appear with this complementizer, and yet they are still described as taking the "finite complement clause" (ibid: 205, 274) rather than a non-finite alternative. In such cases, where we clearly have a dominant pattern alongside a limited number of exceptions, I decided to have a single data point only. The peculiar behaviour of different-subject 'want'-contexts in Noon would be recorded as a distributional restriction on the pattern in question, i.e. the CTP class was marked as being 'partially' rather than 'fully' covered by the finite complement (this will be explained in detail in Chapter 6). Similar situations were encountered in Amele and Karo Batak, to name but two languages of this type.

Taking all of these issues into account, the final database comprises a total of $N_{pat} = 228$ complementation patterns, i.e. slightly more than two types of complement on average for each of the sample languages.[10] In light of the previous discussion, these data points often need to be seen as 'macro-patterns' of complementation or 'construction types'. They constitute abstract patterns of grammatical organization for comparative purposes, and in many (probably most) cases, they are unlikely to have direct psychological reality for the speakers of a given language. Even where these patterns are not conflations of different variants and hence appear to be rather homogeneous form-meaning pairings, one cannot jump to the conclusion that they represent grammatical generalizations that speakers actually entertain. In usage-based models of linguistic knowledge (e.g. Langacker 2000, Tomasello 2003, Bybee 2006, 2010), it is assumed that speakers operate primarily with low-scope, often concrete lexically-specific units that they may or may not ultimately generalize to an abstract morphosyntactic pattern. Therefore, what presents itself as a useful generalization for descriptive or comparative purposes may actually be an agglomeration of many individual constructions that speakers connect to each other only on a very local plane.

---

consideration to begin with. This applied, for instance, to Karo Batak, Fongbe, Mangarayi, Somali, Yuchi and Chimalapa Zoque (cf. §4.7 for further discussion of this point).

[10] Throughout the dissertation, the abbreviation N_{pat} will be used when quantifications are made of individual complementation *patterns*, while N_{lgs} will make reference to a certain number of *languages*. At this point, I will refrain from providing an indication as to which languages contributed more substantially to the sample than others. Since this reflects the nature and organization of complementation systems, this issue will be addressed properly in the analytical chapters to come.

This constructionist perspective has been argued for in a number of studies on complementation in individual languages (e.g. Huang 2003 on Chinese, Cristofaro 2008 on Ancient Greek), and Cristofaro (2008) also provides detailed discussion of the theoretical issues involved (cf. also Chapter 2, fn. 23).

3.3 Storage and analysis of the data

The data for this study were compiled in a digital relational database (cf. Dimitriadis and Musgrave 2009 for databases in linguistics), using *FileMaker Pro 10 Advanced* as a database management system. This database had been designed for the larger project on clause combinations mentioned earlier, which my data on complementation contribute to. Therefore, the complement database is an integral part of that larger data management platform and incorporates information that was gathered in the project context, i.e. also by collaborators. The database was primarily designed for the purposes of systematic data collection, storage and analysis by the project members, to be drawn on in current and future publications; this holds for the larger database as well as for my subsection on complementation. In its current form and design, it is thus not intended as a public database like *WALS*. Taking all of these considerations into account, direct access to the complement data can only be granted to the primary readers of this dissertation, who will find a CD-ROM with the relevant file attached. Statutory copies of the dissertation to be submitted to the *Thüringer Universitäts- und Landesbibliothek* will thus yield to data privacy protection by lacking such a data medium. As a general impression, however, and in order to be able to make sense of the references to the database throughout the dissertation, Material 3 in the Appendix contains screenshots of a sample entry of the database.

In the remainder of this section, I would like to give the primary readers a brief introduction to the design of the database as such and to the basic arrangement of the complement data in the database. Upon opening the file on the CD-ROM, the reader should be automatically directed to a view called "Introduction" (if not, the first menu button on the left can be clicked to get there). The information in the database is organized by functional types of clause combinations, i.e. relative, complement and different semantic kinds of adverbial clauses. These are represented as stable menu buttons on the left-hand side. Each of these rubrics then contains a collection of the relevant data points, i.e. upon clicking the respective menu button, one can scroll one's way through all the records filed in this rubric. In order to avoid disseminating data collected by other project members, I deleted the entries for all clause combinations other than complementation in the CD-ROM version of the database.

The information collected for individual construction types is pooled and displayed in a different format at the level of "languages", which can be accessed by the second menu button on the left. This menu is structured into three tabs, viz. "General information", "Grammatical profile" and "Clause combinations". The first one, illustrated by the first screenshot in Material 3, contains the genealogical and geographical signature of each sample language. The "grammatical profile", displayed in the second screenshot, contains information on basic morphosyntactic parameters that are particularly relevant to the project and the present research. The information is

chiefly organized into categorical variables of constituent order, morphological structure and alignment, which all correspond to individual chapters in *WALS Online* (Dryer and Haspelmath 2011). We have thus greatly profited from the information collected by the *WALS* authors in question, although we investigated (i.e. checked and sometimes modified) every data point again ourselves. Some prose information on these parameters is occasionally recorded in the larger boxes right next to the variables, but an elaborate description of each language, both with respect to its general grammatical profile and its system of clause combinations, was produced in the form of a 'data sheet', i.e. a comprehensive text document for each language. Therefore, the prose information in the database itself is rather sparse, commenting only on some of the coding decisions. The information gathered in this part of the database will be drawn on especially in Chapters 5 and 6 of the dissertation. Finally, the tab labelled "Clause combinations" contains a list of all the individual data points collected for each language, sorted into adverbial, relative and complement clauses. Again, only the fields on complement clauses are filled in on the CD-ROM.

After this general introduction, I would like to give the reader some sense of direction for the "complementation" rubric, which has a rather complex organization (cf. also screenshot III in Material 3). Each entry begins with database-related information, such as the unique "ID" given to each complementation pattern (e.g. "CompAwa1") or an indication of the primary sources of information (in the form of "reference IDs" and page numbers). The first major section of actual data, highlighted in ochre, is concerned with the internal structural properties of each complementation pattern, up to possible diachronic sources of its markers. The variables on the left-hand side will be developed, motivated and discussed in detail in Chapter 4 of the dissertation, so there is no need to go into them here. The adjacent boxes often contain justification of the coding decisions, though again not all of the information gathered on each pattern was copied into the database. Also, since the database is meant for 'internal' use only, the information collected in the prose boxes is not intended to be well-worded or publishable; it is likely to contain many notes (rather than complete sentences), typing errors etc. It is thus a 'workbench' rather than a polished product. Further down (but still in screenshot III), a section coloured in light blue is concerned with the linear order of each complementation pattern relative to its matrix; the variables relate to complements in P/T, S and A function separately and will be analysed in detail for S- and A-clauses in Chapter 5. The remainder of the entry (displayed in screenshot IV) is now split into a left-hand side and a right-hand side. On the left, one can find a record of the functional profile of each complementation pattern, i.e. its co-occurrence possibilities with the various CTP classes defined in §2.5.2. This is first displayed in categorical form, before several prose windows at the bottom (e.g. "Comments", "Productivity", etc.) provide fairly detailed elaboration on the functional profile. All of this information will be exploited in Chapters 6 and 7 later on. Finally, on the right-hand side, one will find an assembly of 'assisting' variables and issues that mostly relate to the integration of the complementation pattern with the grammatical profile of the language as a whole. For example, it is recorded to what extent the complement shows the coding properties as phrasal subjects and objects

("Structural integration with the main clause"), whether there are "placeholders for extraposed subject clauses", to what degree there are "animacy constraints on subjects" in the language, and how different syntactic constructions in the language (e.g. case marking, word order, etc.) define grammatical relations like {SA}, {SP}, etc. ("GR-defining constructions"). The relevance of these variables will become clear in the subsequent chapters. It should be pointed out that the boxes on "Integration with the main clause" and "GR-defining constructions" can basically be glossed over; in comparison to all others, the information contained by them is more preliminary, i.e. not collected as rigorously and not meant for any kind of statistical evaluation; some of the variables were initially intended to be analysed in the context of the dissertation, but were later on discarded and hence not pursued systematically anymore. But since the database is for 'internal' usage only, I have kept the information in the present state. Conversely, some categorical variables that will be used in subsequent chapters cannot be found in that particular form in the database. These were specifically created for the purpose of individual analyses, which were managed and conducted in external spreadsheet and statistical software.[11]

This completes our introductory tour of the database and of the methodological preliminaries more generally. It will have become clear in the course of this chapter that each methodological step involved in the present research – from choosing the languages over identifying and selecting construction types to the specific coding of linguistic variables – was a decision-making process. Decisions to neglect a particular pattern in language X, to split a given construction type into two data points or to conflate two types into one, are bound to be controversial, and I suppose that any typologically informed reader could, in principle, disagree with every single decision. Comparative linguistics is an enterprise that imposes a particular *interpretation* on the descriptive materials available, and I am aware of the fact that broad typological coverage easily runs the risk of overlooking or misclassifying language-specific details. I am taking this risk here for a particular purpose: to make a first move towards an empirically based typological outline of organizational aspects of complementation in the world's languages. Therefore, with regard to most of the questions addressed in the following chapters and the specific statistical methods to be employed, the study can be called exploratory in nature.

---

[11] All statistical analyses to be presented were conducted in the open-source software *R*, version 2.12.0 (*R* Development Core Team 2010).

4

The internal structure of complementation patterns

4.1 Introduction

In view of the cross-linguistic diversity of grammatical techniques for clause combining, a typological comparison of the relevant constructions is a challenging enterprise. The most common solution, also adopted in the present study, is to regard subordinate clauses as complex morphosyntactic entities varying along an array of logically independent parameters. Pioneering papers advocating this approach include Bossong (1979), Haiman and Thompson (1984) and Lehmann (1988), which thus represent specific proposals of what has come to be known as '**multivariate typology**' more generally (Bickel 2007, and cf. Bickel 2010 for an application to clause-linkage). The present chapter pursues such a multivariate approach to the internal structure of complementation patterns. The chapter as a whole provides a foundation for the rest of the dissertation, developing a catalogue of variables which will be used in subsequent chapters to explore cross-linguistic correlations between structural, behavioural and functional aspects of complement clauses. My primary focus in this chapter, therefore, is with the thorough description of the pertinent variables and their illustration from the sample languages, as well as with the analytical problems that some of the attested patterns pose for the distinction between complement clauses and complementation strategies. Since one of the central concerns of later chapters is the integration of subject clauses into the typology of complementation systems, care will be taken in the present chapter to include sufficient examples of subject clauses; the analytical parameters as such are, of course, applicable to complementation patterns in general, regardless of their specific syntactic function in the main clause.

In any typological study, the selection of variables crucially depends on the domain of investigation and the specific research agenda. Bickel's works, for example, often aim at an exhaustive coverage of the morphosyntactic variation that a given phenomenon (e.g. 'nonembedding clause linkage' in Bickel 2010) displays; consequently, the number of variables and their associated levels are determined 'autotypologically' (i.e. with as many fine-grained distinctions as are found in the data,

cf. Bickel and Nichols 2002). In the present study, this exhaustive coverage was not only impossible from a practical point of view[1], but also not intended in the first place. Recall that a central question of the present research is for the semantic and syntactic functions that complementation patterns fulfil in the world's languages, and from these perspectives, it is critical to examine how complement clauses are adapted in their structure to various distributional contexts. In this connection, the key issue is, of course, that of **nominalization**. As particularly Lehmann (2002: 54) points out, complementation always amounts to conceptual nominalization: In prototypical complementation, an action is construed as a referential expression, and this is why complementation structures normally fill argument positions of other predicates. In structural terms, then, complementation patterns are exposed to a particular tension to retain a form that is best suited to rendering their propositional content (i.e. what has been called a 'sentential' or 'verbal' structure) and to adapt, at the same time, to the nominal environments in which they come to be used. In the literature, it has been emphasized repeatedly that the adoption of a nominal distribution often triggers two concomitant but logically distinct types of process, i.e. *de*categorization as the loss of verbal or sentential properties, and *re*categorization as the acquisition of nominal features (Malchukov 2004).[2]

In the present study, the choice of variables for describing the internal structure of complement clauses intends to reflect the workings of these two processes. I will begin by examining the morphological status of the complement predicate (§4.2), i.e. explicit morphological indications on the head of the complement that signal its decategorization and, in most cases, also a certain type of recategorization. Secondly, I will briefly survey patterns of TAM marking and verbal modification from the same perspective (§4.3), before the most substantial section of the chapter details the argument-structural configurations inside the complement along a variety of different parameters (§4.4). In §4.5, I examine what I call the 'flagging' of the complement by various boundary-marking devices, many of which serve the (further) recategorization of the complement as a nominal entity. The chapter ends, in §4.6, with the convergence of the various parameters on a *degree* (or index) of nominality that can be measured for each complementation pattern in the data. This 'measuring approach' will allow us to appreciate differences between relatively similar constructions within the same language (e.g. subtle differences between two types of nominalization), and to compare

---

[1] It is obvious that the whole approach is extremely time-consuming and, given the size of the present sample (N = 100 languages), which goes way beyond Bickel (2010), exhaustive coverage would have exceeded the limits of a single study or researcher. What is more, the sample size would also have precluded retrieving exhaustive information on, say, the scope of illocutionary-force operators, or the specifics of negative marking, in all complementation constructions (some of the sources are entirely reticent about these issues to begin with). Therefore, concentration on a certain number of theoretically most relevant and well-described variables was a sheer necessity. Having said this, it may be noteworthy that, with the exception of illocutionary force and negation, the present catalogue of variables (including the correlational ones of the subsequent chapters) still covers all the analytical dimensions that Dixon (2006a: 21f.) judges relevant to the typological study of complementation patterns.

[2] Two alternative terms that are often cited in this connection are '**deranking**' (Stassen 1985, Cristofaro 2003) and '**desententialization**' (Lehmann 1988). The former term is neutral with respect to whether or not recategorization takes place, i.e. it embraces all kinds of structural deviations from independent clauses. Desententialization, by contrast, is explicitly intended by Lehmann to capture both de- and recategorization processes simultaneously.

across languages the amount of de-/re-categorization that characterizes each complementation system.

4.2 The morphological status of the predicate

As was suggested the introduction above, my discussion of the internal structure of complements will proceed outwardly from the core of the complement proposition, i.e. its predicate and modifiers related to the predicate, to its arguments and finally its external marking. The variable to start with, then, will be what I call the morphological status of the predicate. This term relates to the basic morphological shape of the complement predicate in comparison to that of predicates in independent clauses. Disregarding the inflectional categories of TAM and person indexation (which will be examined as separate variables), the question is thus for any overt morphological indications that decategorize the complement predicate and in this sense change its status from an 'independent verb form' to a 'dependent verb form'.

Independent verb forms, alternatively known as 'balanced' in the literature (Stassen 1985, Cristofaro 2003), are typical of complementation patterns that, as a whole, do not differ from an independent clause. In Begak Ida'an (Western-Malayo Polynesian: Malaysia), for example, "a sentential complement has all the properties of an ordinary sentence. The verb can bear any inflection. The word order of constituents within a sentential complement is subject to the same restrictions as that of a main clause." (Gooudsward 2005: 334) Similarly, complementation in Mandarin Chinese can be rendered by simply inserting an independent clause into the subject or object position of a main clause (Lehmann 2002: 55, Yue 2003: 115):

(34)   Mandarin Chinese (Sino-Tibetan, Sinitic: China; Li and Thompson 1981: 603)
[*Dà   shēng   niàn   kèwén*]   *kěyǐ   bāngzhù   fāyīn.*
 big   voice   read   lesson   can   help   pronunciation
'Reading the lesson aloud can help one's pronunciation.'

What is also subsumed under independent verb forms is a structure known as 'clausal nominalization' in the literature (e.g. Comrie and Thompson 2007: 376). Such constructions are characterized by nominal marking on the clause boundaries and, in some cases, on the internal subject argument, but crucially retain the balanced morphological structure of the verbal head (i.e. there is no overt nominalizing marking on the predicate). This is illustrated in (35) below from Jamul Tiipay:

(35)   Jamul Tiipay (Hokan, Yuman: USA, Mexico; Miller 2001: 224)
Puu-ch   [*nya'wach   neyiw-x*]=*pu*   *map.*
that.one-SBJ   we.SBJ   come.PL-IRR=DEM   want
'He wants us to come over.'

In this example, the object clause is flagged by a demonstrative clitic to indicate its nominal function in the main clause, but the complement is a fully clausal structure with canonical marking of the internal subject and a regularly inflected predicate. In particular, the form of the verb as such is 'independent' in the above sense.

Let us now turn to **dependent verb forms**, which come in a wide variety of different guises across the world's languages. For the purposes of the present study, they have

been organized into five larger groups, viz. nominalizations, converbs, participles, bare verb stems, and 'other' dependent forms. Each of these will be dealt with in turn.

4.2.1 Nominalization

Nominalization as understood here refers to what other researchers have dubbed '**lexical nominalization**' (e.g. Comrie and Thompson 1985|2007), a process that crucially "involves the head, causing a change of its categorical status as compared to independent clauses" (Koptjevskaja-Tamm 1993: 49). This contrasts with the 'clausal nominalizations' from above, which lack this derivational element on the complement verb. A typical example of a lexical nominalization is the Georgian Masdar, illustrated below:

(36) Georgian (Kartvelian: Georgia; Hewitt 1995: 629)

  Ar   še+i+jl+eb+a   [*tbilis-is*   *da-c-em-a*].
  not   it.is.possible   Tbilisi-GEN   PREVRB-fall-TH-NMLZ.NOM
  'It is impossible that Tbilisi fall.'

The defining feature of nominalizations in the sense intended here, then, is that they are productive morphemes for deriving action nominals from verbs, regardless of the morphosyntactic complexity of the resulting structure.[3] As we saw in §2.3.2, 'strongly nominalized' action nominals assume the internal structure of NPs (like the Georgian Masdar above) and are hence regarded complementation strategies, while more verbal types of action nominals are genuine complement clauses. However, this is determined by the properties of other parts of the clause; what is at stake here is exclusively the presence of a nominalizer on the complement verb, and this cuts across strong and weak types of action nominals. In my data, 28.5% of all complementation patterns (N_{pat} = 65/228) exhibit nominalizing morphology on the dependent predicate, making it a prominent grammatical technique in complementation systems. The morphosyntax of such nominalizations has been discussed extremely thoroughly in the literature; relevant references, discussions and many more examples of action nominals will be provided throughout the dissertation and need not be given at this point. Instead, I will focus on a number of complementation patterns in the data whose status as nominalizations is more controversial and hence requires some discussion.

The issue in question revolves around the diachronic and synchronic relationships of nominalizations with two other commonly recognized types of 'non-finite' verb forms, viz. converbs and infinitives. **Converbs**, to be discussed more systematically in §4.2.2 below, can be defined as dependent verb forms "whose main function is to mark adverbial subordination" (Haspelmath 1995: 3), such as a non-finite verb form specialized in the expression of certain temporal relations (e.g. Diyari *thayi-rna* 'eat-

---

[3] Many languages have morphemes dedicated to this function, but sometimes the nominalizers found in complementation are 'participant nominalizers' elsewhere in the grammar. In Huallaga Quechua, for example, the complements of perception verbs are marked by *-q*, which is, at least synchronically, an agent nominalizer (e.g. *pishtaku-q* 'slaughterer', cf. Weber 1989: 53) and the subordinator used in subject-extracting relative clauses. In §7.2, it will be shown in more detail that systematic overlaps between 'process' and specific types of 'participant' readings of nominalizations play an important role in the diachronic development of complementation constructions.

CONV = 'while eating' (ibid.)). The converbal morphemes themselves have thus been argued to be functionally parallel to nominalizers, in the sense of creating 'verbal adverbs' (Haspelmath 1995: 4; but cf. Ylikoski 2003 for a detailed critique of this approach). **Infinitives**, by contrast, have a less clearly defined profile. In Noonan (2007: 67), they are regarded as a distinct "morphological type of complementation" (contrasting with sentence-like, nominalized and participial complements). However, his definition of infinitives does not actually relate to the morphological shape of the complement predicate, but to the requirement that they "do not bear syntactic relations to their notional subjects" (ibid.), and to their otherwise clause-like internal structure. This poses the question of how his infinitival complements should be treated in regard to the form of their verb. Closer inspection of the grammatical descriptions of our sample languages reveals that this needs to be decided on a case-by-case basis, since the term 'infinitive' is used very heterogeneously in the sources.[4] And such an individual examination of the allegedly 'infinitival' constructions in the data soon discloses their intimate connection with nominalizations and converbs.

The problem is, of course, that the verb forms commonly labelled 'infinitive' and those called 'converbs' both "arise diachronically from adpositional or case forms of verbal nouns", i.e. of lexical nominalizations in the above sense, and that there is "a continuum of grammaticalization" between them (Haspelmath 1995: 28): specific case-forms of verbal nouns first come to fulfil adverbial functions (in the relevant instances typically purposive functions), and once grammaticalized as new converbal markers of this domain, they come to serve as complements of semantically compatible CTPs, thereby returning to the nominal functional domain from which they had originated. This 'cyclic' development has been established as a cross-linguistically valid type of diachronic change (Disterheft 1980, Haspelmath 1989), and we will have more to say about it in §7.2. The general point, however, is that because of the cyclic nature of the grammaticalization process, the morphology of the complement predicate is still, at least in part, that of a nominalization (as Haspelmath 1989 intends to show). It is only because (i) relatively few morphosyntactic traces of the original nominalization construction are left, (ii) the construction is functionally firmly established in the domain of complementation, that it is seen as a distinct form, called 'infinitive'. Incidentally, the same synchronic properties also hold for so-called infinitives that are simply more 'verbal variants' of action nouns (without necessarily involving converbal diachrony), as in Noon and other African languages in the sample (e.g. Hausa, Krongo):

(37) Noon (Niger-Congo, Atlantic: Senegal; Soukka 2000: 274, 254)

    a. *Mi hot-in* [*ki-hay-kaa ƴaal-aa*].
        1SG see-PRFV   INF-come-DEF man-DEF
        'I saw the man come.' (lit. 'I saw the man's arrival.')

---

[4] This has also been noted by other typologists, including Noonan (2007: 67) himself. Dixon, in his approach to the typology of complementation, thus recommends to "shun the term" altogether (Dixon 2006a: 44). As will be seen shortly, this is precisely what I intend to do in the coding of the morphological form of the verb.

b. [*Ki-lín   hareen*]   *mesk-in.*
    INF-cultivate peanuts   be.hard-PRFV
    'To cultivate peanuts is hard.'

As can be seen, both examples contain the same 'Infinitival' morpheme; however, in (37a), the *ki*-marked predicate functions "as the head of an NP", while it is "the head of a VP" in (37b) (Soukka 2000: 175): the VP structure has fully verbal internal syntax, while the NP structure marks its subject in the Genitive and the predicate itself can carry a Definiteness marker.[5]

In sum, the general argument is that there is no cross-linguistically applicable *morphological* concept of 'infinitive' that is distinct from either converbal or nominalizing morphemes. For this reason, I subsumed the so-called infinitives in grammatical descriptions under either the former or the latter label. This was possible (in most cases) because information on the diachronic source and the functional distribution of each complementation pattern in the database had been elicited for the study, anyway, and so this information could be brought to bear on the morphological status of the complement predicate. Specifically, I decided to treat infinitives (and constructions with similar properties that were not explicitly called infinitives) as instances of nominalizations

- if the infinitival morpheme can be related to a nominal form, either etymologically as part of its morphology or synchronically as part of the system of nominal morphology (e.g. if it is a derivational marker outside of complementation, a noun class or case marker, etc.)

- *and* if it is synchronically a firmly established complementation pattern and not an adverbial clause used sporadically for complementation.

The latter criterion also features in the definitions of infinitives in Haspelmath (1995), Nedjalkov (1998) and Ylikoski (2003) and is meant to separate them from converbs, i.e. forms that primarily function in the adverbial domain. These are adjunctive in their syntax (i.e. non-obligatory elements) and have not penetrated far into the complementation system (i.e. they have not been analogically extended to a wide range of CTP classes). The former criterion is meant to separate the relevant constructions from other kinds of dependent verb form more generally. Let me illustrate the workings of the two criteria in relation to a few selected data points from the sample, in the form of a 'fast-track' list: The so-called "Infinitive" in Warao (isolate: Venezuela) is a purposive form that shows no traces of a nominal structure, function or etymology; its origins are probably rather to be found in an intentional mood marker plus manner

---

[5] In this and similar examples from other languages (e.g. Hausa, Krongo, etc.), it is unclear whether a converbal stage was involved at all in the development of the 'infinitival' structure, or whether the latter arose by 'verbalization' (or 'gerundialization') of the action nominal (a pathway described extensively by Fanego (2004) for English; cf. also §7.2). While it is true that all of the infinitival constructions in Noon, Hausa and other languages can be used in purpose clauses, these are restricted to so-called 'motion-cum-purpose' environments (Schmidtke-Bode 2009: 94), and in such cases, the infinitive is often felt to be obligatory (and hence a complement to begin with) rather than a genuine adjunct. It is thus hard to tell whether a genuinely converbal stage was involved diachronically. Interestingly, Ylikoski (2003) defines infinitives precisely as non-finite forms that function as obligatory elements in a sentence, i.e. either as complement or "argumental adverbial", in order in include motion-cum-purpose clauses.

affix, and it shows very limited complementation functions (Romero-Figeroa 1997: 18). It was thus classified as a converb. The "Supine" form in Kolyma Yukaghir (Yukaghir: Russia) originated in the Possessive Dative form of the Action Nominalizer, in which it came to be a purposive converb. In the process, it also developed some formal idiosyncrasies that do not apply to regular forms of the Action Nominalizer (Maslova 2003: 151). Synchronically, this grammaticalized form is firmly established as a purpose clause and has been extended to only few complement-taking predicates. Hence, it primarily fails on our second criterion and was treated as a converb. For a contrast, a few nominalizations: The Serbo-Croatian Infinitive morpheme -ti is synchronically distinct from the Masdar, but it was "the dative form of a deverbal nominal stem in *-ti (Joseph 1983: 102) in Proto-Slavic, and is synchronically widely applicable in complementation, while its purposive use is restricted to motion-cum-purpose. In Urarina (isolate: Peru), the so-called Infinitive has no role to play at all in purposive (or other adverbial) clauses, and it is plausibly related to a derivational nominalizer (Olawsky 2006: 761). This would appear to fulfil both of the above criteria for nominalization. Similar (though of course not identical) arguments apply to so-called Infinitives in Awa Pit, German, Hungarian, Lezgian, Mayogo, Dolakha Newar, Purépecha, Wappo and Wolaytta. One case is left as sitting in between converbs and nominalizations in the above sense: An infinitive-like construction in Malayalam fails on criterion 1 (i.e. it cannot be transparently related to nominal morphology) but is quite productive as a complement clause. In the database, it has provisionally received a 'double coding', but for statistical purposes, it will later be subsumed under nominalizations.[6]

4.2.2 Converbs

A definition of converbs, taken from Haspelmath's (1995) seminal paper on the topic, was given in the preceding discussion (cf. also Tikkanen 2001 or Nedjalkov 1998 for similar characterizations). In accordance with that, patterns coded as converbs in my data can all be argued to involve a dependent verb form that contains a marker clearly dedicated to one or more adverbial functions, and whose uses in complementation are secondary from both a synchronic and (very likely) a diachronic point of view. This applies, first and foremost, to a number of dedicated purpose markers again. In Rama (Chibchan: Nicaragua), there two converbal suffixes for the expression of purpose, -kama and -bang, and both can be used, for example, in desiderative, jussive and similarly 'prospective' complementation contexts (e.g. 'remember to do something') (Grinevald 1990: 222). Purposive converbs with similar properties are found, for

---

[6] For the sake of completeness (and in order to avoid misunderstandings with regard to the coding of specific constructions in the database), let me also point out the following: Complementation patterns whose structure transparently involves a lexical nominalizer and a purposive or other subordinator as two separate morphemes did not enter the whole infinitival discussion to begin with. These clearly involve nominalized verb forms, regardless of their history or their associated synchronic functions (e.g. certain complements in Amele, Imonda, Huallaga Quechua or Tzutujil). In other words, the 'infinitive problem' only arises in regard to monomorphemic markers whose status as either converb or nominalizer needs to be examined more closely, and it is only these cases that the foregoing discussion was about.

example, in Evenki (*-dA*, Nedjalkov 1997: 27), Martuthunira (different-subject purpose inflection, Dench 1995: 256) and even more clearly in Tariana (where the uses in complementation are extremely restricted: the purposive form in *-ka* combines only with one mildly jussive verb ('ask to do something') and with lexically-specific instances of non-canonical S- and A-clauses; Aikhenvald 2003: 551).

But apart from purpose, converbs may also be associated with other adverbial domains, although this is a rather infrequent case in the data. We find it in Matsés, where the converbal forms for 'while' and 'when' have started to encroach on the domain of complementation, as discussed insightfully by Fleck (2006). (38) shows the typical adverbial usage first, followed by an example where the converb is used in the direct-object position of the CTP 'see'; as no NP object is allowed to be present in this construction, the converbal clause acts like a proper complement clause:

(38) Matsés (Panoan: Brazil, Peru; Fleck 2003: 1101 and Fleck 2006: 240)

    a. [*Uënës-bud-sho*]   *matses-n*   *tabote* *dë-bed-quid.*
       die-DUR-**when**   Matses-ERG   torch   tip-tap.away.ashes-HAB
       'As [the torch] starts to die out, the Matses tap away the ashes from the tip.'

    b. [*Mimbi debi-Ø*   *kues-sho*]   *is-o-mbi.*
       2ERG   Davy-ABS   hit-**when**   see-PST-1A
       'I saw you (as you) hit Davy.'

In Awa Pit (Barbacoan: Ecuador, Colombia), a semantically versatile converb construction (used as a linkage device for evoking temporal, causal and conditional relations) has come to be a complement clause for indirect speech (and only this complement function); the latter construction will concern us again later on in this chapter and will be exemplified there (cf. (51)).

In total, there are 10 complementation patterns coded as containing a converbal verb form in the data (10/228 ≈ 4%), plus the Malayalam construction discussed in the previous section as falling in between nominalizations and converbs. As a whole, the converbal technique in complementation has a certain bias towards South American and Eurasian languages in the sample, but given its limited overall number, this cannot be taken to be a reliable areal generalization (cf. Tikkanen 2001 for some distributional observations on converbs in general). Finally, it may be worth pointing out that the converbal status of the verb form, and hence the primarily adjunctive nature of the pattern in question, does not entail that we are automatically dealing with a complementation strategy. If there is sufficient evidence that the converbal structure comes to fill an argument position of a predicate, as in (38b) from Matsés above, it has to be regarded as a genuine complement clause (for at least these CTPs). In my sample, all converbs except for the one in Martuthunira have reached complement-clause status with at least a subset of the CTPs they can attach to: They can all function as non-subject arguments, and in Rama and Tariana, they are also found as subject clauses (though with certain restrictions).

4.2.3 Participles

Besides nominalizations, infinitives and converbs, **participles** have been considered a major type of dependent verb form in the literature. They are usually characterized as "verbal adjectives" (Haspelmath 1995: 2), in the sense that their morphosyntax and primary synchronic function are akin to that of attributive adjectives. With regard to complex sentences, this means that participles are verb forms primarily associated with the domain of attributive noun-modification, i.e. relativization (thus cf., e.g., Lehmann 1984 for participial relative clauses). In this respect, participles parallel converbs, whose primary function is also outside of the domain of complementation. And since they are not normally referential expressions, unlike nominalizations, "the role of participles in complementation […] is limited" (Noonan 2007: 72), a fate they share with converbs. Thus only 2.2% of my data (N_{pat} = 5/228) contain verb forms that could be labelled participles. The host languages are Evenki, Ma'di, Tümpisa Shoshone and – again – Martuthunira and Lezgian. Within this group, the clearest case to match the definition of participles is a construction in Lezgian. In this language, relative clauses are formed by adding one of two suffixes to the verb stem, creating 'Participles' and 'Aorist Participles', respectively. This is illustrated in (39) below:

(39) Lezgian (Nakh-Daghestanian, Lezgic: Azerbaijan, Russia; Haspelmath 1993: 341)

[*Mu'minat-a*   ø   *ktab*   *ga-ji*]   *ruš-a*   *qʰfe-na.*
 Muminat-ERG ø.DAT book   give-**AOR.PTCP** girl-ERG go.away-AOR
'The girl to whom Mu'minat gave the book went away.'

Now, this participial construction can also be used with referential functions, i.e. headless relative clauses and clausal complements, but then it needs to be suffixed with a "substantivizer":

(40) Lezgian (Nakh-Daghestanian, Lezgic: Azerbaijan, Russia; Haspelmath 1993: 365)

Wiri *ajal-r.i-z*   [*Musaq'.a-n* *dide*   *q'e-nwa-j-di*]   *či-zwa-j.*
all   child-PL-DAT Musaq-GEN mother die-PRF-**PTCP-SBST** know-IPFV-PST
'All the children knew that Musaq's mother had died.'

This constraint makes it clear that the participial form itself is not a referential expression, but only capable of modifying a noun; this, and the fact that its uses in complementation are clearly restricted, makes the label 'participle' appropriate.

In all other cases, however, it is only a synchronically strong bond with relativization that has led me to consider the verb forms participial in nature. Under close scrutiny, these markers may turn out to be nominalizers, after all. What they all have in common is that, unlike the genuine participles in Lezgian, they can actually be used as referential expressions themselves: they can function as a headless relative clause and also as a complement:

(41) Ma'di (Nilo-Saharan, Moru-Ma'di: Uganda, Sudan; Blackings and Fabb 2003)

a. *Àràbɨ̀à* [*ópɨ́* *ʔà* *dʒɨ̄-lɛ́*]   *rɨ̀*   *pá* *nā* *ādī* *rá.*
    car   Opi POSS (L-)take-**SUB** DEF leg AFR deflate AFF
    'The car which Opi took certainly has a flat tyre.' (200)

b. *ópí̵   ē-dʒé*   [`ɲā-lɛ́   rá]   rɨ̵   ʔɨ̵.
   Opi   (3)-VENT-buy   L-eat-**SUB**   AFF   DEF   FOC
   'Opi bought the one which was certainly eaten (by someone).' (191)

c. *Má   lɛ̀-ā*   [*èɓi   `ɲā-lɛ́*]   *rá.*
   1SG   (L-)want-OBJ   [fish   L-eat-**SUB**]   AFF
   'I certainly want (me, someone) to eat fish.' (203)

Exactly the same pattern of multifunctionality is attested for the other 'participial' verb forms in the sample. It is thus possible that they are, in fact, nominalizers at least from a diachronic point of view. For Ma'di, Andersen (1996: 295) maintains that "relative clauses are nominalized clauses", although no diachronic evidence is provided to substantiate this claim. Similarly, Evenki complement clauses are said, by Nedjalkov (1997: 23), to make use of a set of "participles", but at least some of the relevant forms may turn out to be nominalizers (Andrej Malchukov, p.c., and cf. Grenoble 2012 and Koptjevskaja-Tamm 1993 for an analysis in terms of nominalization for some of the markers). From this perspective, the evidence for truly participial complements in Evenki is definitely weaker than in, say, Lezgian. Furthermore, the so-called participles are the major complementation technique in the language, being highly productive rather than restricted as complements. There is thus no striking functional gravitation towards relative clauses, which would be needed for genuine participles. Accordingly, the forms in Evenki are coded as intermediate between nominalizations and participles (much like the mixed nominalization/converb case above), and subsumed under nominalizations for statistical purposes. For the remaining languages, there were sufficient reasons to adopt a participial analysis. Thus in Martuthunira, the complements in question were explicitly described as "present relative clauses" (Dench 1995: 255), and these contain specific non-finite verb forms which must hence be seen as participles. In Tümpisa Shoshone and Ma'di, what tipped the scales was the functional bias of the construction towards relativization or, conversely, their limited use in complementation. Ma'di, for example, has two highly productive complementation patterns, while the 'participles' co-occur with "only a few lexical verbs" (Blackings and Fabb 2003: 420).

4.2.4 Bare verb stems

A morphological entity somewhat different from the traditionally recognized dependent verb forms is what I call '**bare verb stems**' here. This is basically a subtractive morphological category, a form of the verb specific to dependent clauses which remains after all (or most) possible morphological marking is stripped away. For so-called 'embedded' and 'merged' complements in Sanuma, for example, it is characteristic that "in the verbs of these clauses there occurs only the stem of the verb (or the stem plus a single-vowel aspect marker)" (Borgman 1990: 104):

(42)   Sanuma (Yanomam: Brazil, Venezuela; Borgman 1990: 104)

   Sa   [*tu   wehe*]   *ta-a   ko-ta.*
   1SG   liquid   dry   see-DUR   return-EXT
   'I (am about to) see the water dry up again.'

One may argue that a distinct category 'bare verb stem' is superfluous because it could be seen as a combination of our other variables: On this line of reasoning, stems would be 'independent', i.e. derivationally unchanged, verb forms with TAM- and/or indexation reduction. However, one has to keep in mind that they clearly constitute dependent verb forms in the respective languages and thus should not, for cross-linguistic analyses involving the verb-parameter, be binned with truly independent predicates. In some languages, there is, in fact, some evidence that the verb form in question behaves differently from an independent one and should thus be seen as dependent. In Lakota, for instance, the so-called 'Infinitive' corresponds morphologically to the third-person indicative form of the verb (because in this inflectional form, subject and object indexes are zero-morphemes). However, when it functions in dependent clauses, this verb form can actually be used in contexts where the referent of the (implicit) subject is not a third-person referent, as in (43) below:

> (43) Lakota (Siouan: USA; Buechel 1939: 301)
>
> [*U*]   *ŭyaśipi.*
> come.INF   you.tell.PST.us.it
> 'You told us to come.'

Therefore, the 'Infinitive' is a grammaticalized, or 'frozen' form, of an erstwhile independent verb and thus has a different synchronic status. This is reflected by putting it into the category 'bare stem'.

Overall, there are 7 instances of such stems in the data (= 3%). Apart from Sanuma and Lakota, they are found in Barasano, Choctaw, Gooniyandi, Koyra Chiini and Semelai. In Choctaw, the situation is very similar to Lakota, the dependent form simply being the "citation form" of the verb but with idiosyncratic features (Broadwell 2006: 281). In Gooniyandi and Koyra Chiini, the stem is accompanied by an overt subordinator, but this is a clitic in Gooniyandi (McGregor 1990: 400) and a free 'Infinitival' morpheme in Koyra Chiini (Heath 1999: 304); the markers are hence not part of the morphological paradigm of the verb and, accordingly, cannot be counted as nominalizers, converbs or participles (cf. Haspelmath 1995 for this criterion).

4.2.5 Other dependent forms

Finally, there are verb forms in the data that do not readily fit into the categories established so far and will, therefore, collectively be referred to as '**other dependent verb forms**'. Some of the relevant cases may well turn out to be instances of nominalizations, participles, etc., but in the absence of clear evidence to this effect, I have opted for a more conservative coding policy which keeps them separate for now. An example comes from Jarawara (Arauan: Brazil). In this language, the predicates of complement clauses are marked by a distinctive phonological change, the verb-final vowel raising from /a/ to /i/ (Dixon 2004: 91). Crucially, while this is very similar to a productive morphological process for deriving deverbal nouns in the language (/a/ > /e/), and the complement clause itself has a number of nominal properties, Dixon (2004, 2006b) does not refer to nominalization in his description of the construction. He rather explicitly contrasts them with "nominalized clauses" (ibid.: 482), which exist

as a separate construction type. In view of these descriptive details, I have coded the complement clause as containing an 'other' dependent verb form.

A similar process of phonological change, this time truly unrelated to nominalization, occurs in Wappo. In this language, subordinate clauses of all kinds (i.e. not just complement clauses), are characterized by dropping the glottal stop as the final phonological segment of the subordinate predicate. This "can be seen in the verb form *hak'še* in the example below, whose form in independent clauses would be *hak'šeʔ*" (Thompson et al. 2006: 109):

(44) Wappo (Wappo-Yukian: USA; Thompson et al. 2006: 109)

Ah	[*ce*	*k'ew*	*i-ø*	*hak'-še*]	*haṭis-khiʔ.*
1SG.NOM	DEM	man	1SG-ACC	like-DUR.**DEP**	know-STAT

'I know that the man likes me.'

As this phonological modification is characteristic of subordinate clauses, it is explicitly argued to constitute "a non-finite, or dependent (DEP), verb form" (ibid.). And since the occurrence of these forms is not biased towards a specific type of subordinate clause, I refrain from coding it as participial, converbal or nominalizing, opting for 'other dependent form' instead.

An altogether different type of dependent verb form is created when predicates in certain subordinate clauses must be deprived of an otherwise obligatory illocutionary-force ('sentence-mood') inflection, while other inflectional categories (person marking, TAM) may be retained and vary independently with the specific functional domain in which the subordinate clause is used. An example of this pattern is found in the Austro-Asiatic language Santali. There are several distinct complementation techniques in this language, and in some of them, the verb is distinguished from an independent sentence by the absence of the "indicative marker *-a*" (Neukom 2001: 181), this being the major feature to discriminate paratactic and hypotactic constructions in the language. The complementation patterns affected by this kind of dependent marking may then contrast again in regard to their TAM inflections, which thus need to be coded independently. (45) below illustrates an illocutionary-force-reduced complement with very few other restrictions:

(45) Santali (Austro-Asiatic, Munda: India; Neukom 2001: 181)

[*Din-din*	*sɛn-a-ko-dɔ*]	*baŋ*	*boge-a.*
day-REDUP	go-APPL-3PL.OBJ-TOP	NEG	good-IND

'It's not good to go to people (lit. 'them') every day.'

A similar pattern is also attested for so-called 'dependent final clauses' in Menya (Trans-New Guinea, Angan: Papua New Guinea). More generally, it should be pointed out illocutionary-force reductions were only taken into account and coded as 'other' forms (i) if this relates to an obligatory inflectional category of the complement verb and hence affects its morphological shape; (ii) if the verb in question is not already characterized by more specific dependent morphology, such as nominalizations, participles, etc. In Korean, for instance, where certain types of complement clauses lack the otherwise obligatory "sentence enders" (Sohn 1994: 53), there was no need to

assume 'other dependent verb forms' since the predicates in question are clearly nominalized.

The remaining cases of 'other dependent verb form', which add up to a total of this category of 7% of the data (N_{pat} = 16/228), are found in Burmese, West Greenlandic, Lango, Malayalam, Purépecha, Somali, Wari' and Wolaytta. In all of those cases (except for Lango), we are again dealing with a relatively 'finite' verb form that cannot easily be subsumed under nominalizations, participles or converbs. In some cases, e.g. Wolaytta, there is considerable discussion in the literature on the most suitable analysis of the forms in question (cf. Wakasa 2008), so my label 'other' acknowledges such controversies. There is no room to present each of the above cases individually here; readers specifically interested in this issue are kindly referred to the database, where individual analyses have usually been given justification.

4.2.6 Distributional aspects of the different verb forms

As a summary of the preceding sections, Table 3 lists the various verb forms in complementation patterns again, along with an indication of their occurrence frequencies in the present sample:

Table 3. Morphological status of verb forms in complementation

Verb form	F(abs)	F(rel)
Independent	127	55.7
Nominalization	65	28.5
Converb	10	4.4
Participle	3	1.3
Bare verb stem	7	3.1
Other dependent form	16	7.0
Total	228	100

The distribution of these verb forms becomes more interesting if we investigate how the individual categories combine in the complementation systems of entire languages. Most complementation systems in the data (N_{lgs} = 55/100) exhibit a combination of at least one independent verb form and at least one dependent verb form. This is the most robust type of system to the extent that it is found in all macro-areas, and quite prominently so. It is particularly distinctive of the languages of Africa (13/17 systems = 77%), Eurasia (11/15 = 73%) and South America (14/23 = 61%). However, one may have expected the overall percentage of mixed systems to be much higher; in fact, Noonan (2007: 146) even predicts that *all* languages will have a sentence-like indicative complement type and some sort of reduced complement type in opposition to that. This claim can, I believe, be upheld, but only (i) if one adopts as broad a definition of complementation as Noonan does, and (ii) if the statement is interpreted as including structural variables other than the verb form as such.

As for the first point, we have to recall from §2.1 that Noonan defines complementation entirely in semantic terms. If a stricter delimitation of the domain is opted for, as in Dixon (2006a) and in the present study, it is possible that some of the allegedly mixed dependent-independent languages turn out to be different, gravitating

towards either dependent or independent verb forms only. In other words, the distribution in the present study is likely to be explicable by the specific selection and classification processes that I adopted. This can be seen most clearly when we examine those languages in my data that appear not to make use of independent verb forms, i.e. to lack a sentential type of complement. This applies to 14 languages overall, which are fairly evenly distributed from an areal perspective (with roughly 2 relevant data points per macro-area). In one subgroup of those languages (N_{lgs} = 8), there is unambiguous evidence for a sentential structure that would count as a complement in Noonan (2007), but not in more restrictive approaches. The 8 structures in question all occur in direct-speech constructions, and following Dixon (2006a) and others, these should only count as part of a complementation system if there is some evidence for their argument status (or embedding in the main clause). This is not the case, for example, in Kwazá, Lavukaleve, Warao, Wari' and Yimas, where we often find explicit argumentation to the effect that the quotative clause forms an independent discourse and is not syntactically part of another clause. In yet other languages (e.g. Evenki, Dolakha Newar, Huallaga Quechua), direct-speech *can* be embedded as an argument, but there already is an indirect-speech construction in the system (which uses a dependent verb form) and hence the direct-speech construction was not taken into account in addition to that (cf. §2.5.2 again for this decision). The result is that most of the 8 languages above rely on dependent verb forms in my data, and I believe that this captures the complementation systems in question, and hence also the general grammatical 'character' of the language, quite faithfully. For example, Terrill (2003) repeatedly emphasizes for Lavukaleve that "nominalisation is a common and productive process in [the language]" (Terrill 2003: 347), whereas "there are no complement clauses" (ibid.: 351). Similar statements on the strong or even exclusive reliance of complementation on non-finite verb forms are also available for Evenki, Kolyma Yukaghir and Yimas.

The other subgroup of languages with exclusively dependent verb forms in my data (N_{lgs} = 6) can at least partially also be explained by interpreting Noonan's wording in specific ways. In my data, Purépecha, Somali, Wappo, Yukaghir, West Greenlandic and Wolaytta do not have a fully sentential type of complement. It is conspicuous, however, that all of them (except for Yukaghir) exhibit a verb form of the 'other' type alongside a nominalization. As was mentioned above, such 'other' dependent forms often have a rather finite appearance, especially in comparison to the clearly nominalized construction they contrast with in their respective system. Therefore, these 6 complementation systems also adhere to Noonan's generalization in being more "sentence-*like*" than the clearly "reduced" complement they contrast with. In my analysis, however, all of the patterns in question involve some significant changes to the morphological form of the predicate, and so I coded them as having dependent verb forms.

In relation to the second point from above, we can turn to the opposite end of the spectrum, i.e. the sizeable number of complementation systems in my data that do not appear to have dependent verb forms (N_{lgs} = 31/100). This figure, as it stands, is misleading in at least two ways. First, a caveat of the present study that has to be

recalled from §3.2 (esp. fn. 9) is that for some of the sample languages, a nominalization construction is known to exist (and likely to have some (though maybe restricted) function in complementation), but the information on this pattern was not sufficient to take it on board as a data point. In reality, then, some of the systems with independent verb forms only may turn out to be mixed systems, after all (e.g. Fongbe or Mangarayi), at least if their dependent constructions can fulfil the definitional and functional requirements imposed upon all constructions in my selection process. Second, and more importantly, the fact that many constructions in the data have morphologically independent verb forms does not entail that the overall structure of the complement is fully sentential. A well-known case is Japanese, where both *no-* and *koto*-complements have synchronically independent verb forms, but allow for Genitival coding of their subjects and are flagged, at the clause boundary, by a free nominalizer (e.g. a grammaticalized noun). Therefore, they are less sentential than a third type of complement (the so-called *to*-clauses), despite the fact that all three patterns display independent verb forms. In view of this, Japanese, too, turns out to fall under Noonan's generalization, and it underlines the necessity of a multivariate approach to complementation patterns. A full investigation of *degrees* of desententialization will, therefore, have to wait until the end of the present chapter.

In closing this section, it should be pointed out, though, that there *are* complementation systems in the data in which lexical nominalization plays a marginal or even no role at all. Thus Fehan Tetun distinguishes 'sentential' and 'reduced' complements, but the reduced ones do not involve nominalization. Moreover, van Klinken (1999: 278) explicitly points out that "there is no[t even] evidence for nominalisation as a complementation strategy in Tetun". In much the same way, some other languages of the sample have nominalization as a productive derivational process in the morphology, but do not employ it at all or only in a restricted way in complementation. In Begak Ida'an (Austronesian: Western-Malayo Polynesian, Borneo: Malaysia), for instance, nominalizations can function as expressions of manner ('way of doing something'), but are not used as complements of our CTP classes (Goudswaard 2005: 252). Similarly, Kayardild (Australian, Tangkic: Australia) has a well-developed system of nominalizations, but these are simply not productive in argument functions and thus of rather little importance to the domain of complementation as defined here (Evans 1995: 270).[7] In Barbareño Chumash (Chumash: USA), so-called 'nominalized clauses' are extremely productive, but they are chiefly built around participant nominalizers and thus do not normally encode genuine process complements; the latter are rendered, and also presented in the grammar, as a completely distinct system (Wash 2001: 89f.), and this is based on more sentential constructions, i.e. independent verb forms in my coding schema. These cases will suffice to show that while nominalization is probably a universal derivational process, it need not be exploited for complementation.

---

[7] The only exception here is their function as an adjoined complementation strategy for perception verbs, but this function is productively covered by the major complementation pattern in the data. Hence, the nominalized pattern was not taken into account as a separate data point.

4.2.7 Theoretical aspects of the different verb forms

As a final step, we are now going to evaluate the different types of verb form from the overarching perspective of this chapter, viz. the issues of de-/re-categorization. So-called 'independent' verb forms do not involve any decategorization of the predicate since they amount to the same morphological form of the verbal base that would also be found in independent clauses. Nominalizations, converbs, participles and bare stems, on the other hand, are clear signals of decategorization. It is arguably true that, within this group, nominalizations stick out again because they also indicate the recategorization of the verbal base as a nominal entity. By contrast, the potential of participles and converbs to act as category-changing, or similarly "**transcategorial**" (Malchukov 2004), devices is more contested in the literature. In other words, it has been debated to what extent the terminology of "verbal adjective" and "verbal adverb" is really justified (cf. Ylikoski 2003 for a detailed paper devoted to this issue). However, for some of the participles and converbs discussed above, it was shown that they are likely to involve nominalization at least etymologically, and that some of the participles, in particular, are fully capable of constituting a referential expression by themselves. (Recall that it was mainly their synchronic distribution that made us call them participles). In view of this situation, it is problematic to clearly differentiate between nominalizations, participles and converbs as far as their force of recategorization is concerned. For statistical purposes, they will thus be treated, along with bare stems, as a group of verb forms that maximally contrasts with that of independent forms (cf. further below).

The class of 'other dependent forms' is difficult to evaluate because of its heterogeneity. As was mentioned in §4.2.5, the dependent forms in Jarawara, Wappo and Lango can safely be called decategorizing devices. All others, by contrast, fall somewhere in between the two opposing poles from above; they have not been susceptible to an analysis in terms of the traditional dependent verb forms (i.e. none of them has been described in the materials as participial or as a verbal noun), but they still differ visibly from the verb forms used in independent clauses in the respective languages. If decategorization is seen as a continuum, as in Lehmann (1988), then these remaining forms are less strongly decategorized than the traditionally assumed dependent verb forms.

The recognition of continua of grammatical coding is important because one of the ultimate goals of the present chapter was to develop a way of *measuring* the degree of decategorization for each complementation pattern. The idea is that every data point can be allocated a **cumulative index** of de-/re-categorization, to which each variable in the catalogue contributes a numerical value. I am, of course, aware of the fact that the translation of categorical data into measurements (i.e. ordinal- or interval-scaled data) is generally a problematic step. However, the very nature of complementation, and of the variables that are used to define complementation constructions, is not merely categorical but inherently scalar, i.e. directed at fully clausal expression formats on one end, and at fully nominal ones on the other. Put differently, there is an inherent logic in the grammar of complementation that is structured around *descreasing* amounts of sententiality and *increasing* amounts of nominality. From this angle, I conceive of this

dissertation as a possibility to explore the potential of a scalar approach to data on complementation.[8] Keeping this agenda in mind, the above argumentation would suggest the following **numerical coding** of the verb form in complements (Table 4):

Table 4. Numerical coding schema: Degree of decategorization of complement verb forms

1	Nominalizations, participles, converbs and bare stems (= highly decategorizing and at least partially also recategorizing)
0.5	'Other' dependent forms (except for the above-mentioned cases)
0	Independent verb forms (= no de-/re-categorization)

In the same spirit, we are now going to move on to the other variables in my data, beginning with two aspects of what may be called 'predicate modification'.

4.3 Predicate modification

4.3.1 Tense-aspect-mood (TAM) inflection

The rendition of time, aspectual structure and modality in complementation is a thoroughly studied area (cf. already Ransom 1986 for a summary of classic research), and no attempt will be made here to provide a qualitative outline of the manifold expression types that these categories yield across the world's languages. Instead, I will confine myself to a survey of how TAM has been parameterized in the present study in order to qualify as a variable whose association with other parameters of complementation can be measured.

In cross-linguistic surveys and large-scale analyses of complementation (e.g. Ransom 1986, Noonan 2007), TAM is usually approached from the perspective of the complement-taking predicates or the semantics of the complementation construction as a whole: Some complements are thus said to have their time reference 'determined' by their CTP (e.g., desiderative complements are future-oriented relative to the time of the desire), while other CTPs allow for 'independent time reference' (e.g., the temporal reference of a quoted utterance is completely 'independent' of the time at which the quoting takes place), etc. In this study, we will approach the issue from the perspective of the complementation pattern, and mainly from a structural perspective: Keeping the processes of de/re-categorization in mind, we can ask ourselves if and how the inflectional categories relating to TAM are expressed on the complement verb. This is similar in spirit to Cristofaro's (2003) approach, but makes some distinctions that are specifically geared towards complementation.

Before we begin, however, a brief note is in order on the way in which 'TAM' is understood for the present purposes. To start with, TAM is seen, somewhat in contrast to the creed of multivariate typology, as a single parameter of investigation. This may not be uncontroversial, but it should be kept in mind that I am not interested here in the specifics of tense, aspect and modal marking in complementation, but rather in whether this dimension of predicate specification, as a whole, needs to be severely

---

[8] Note also that exploratory analyses of grammatical data in other areas of comparative linguistics have followed the same procedures, e.g. Szmrecsanyi and Kortmann (2009) on varieties of English.

reduced or modified (resulting in signals of decategorization) or whether it continues to be expressed as in independent clauses. Furthermore, since none of the categories T, A or M is normally associated with recategorization (i.e. they cannot be transformed into corresponding nominal categories), unlike most other variables in this chapter, their individual expressions are less relevant in the present context; for this reason, too, I decided to treat them as a single parameter.

As far as the operationalization of this parameter is concerned, it must be mentioned that TAM is understood primarily as an inflectional category of the complement predicate and thus refers to what Functional Grammar calls 'operators' rather than 'satellites' (e.g. Hengeveld 1989). Satellite expressions of TAM in the form of temporal adverbs etc. were not normally taken into account.[9] As a further restriction, the term 'mood' is used almost exclusively in reference to modal values such as (ir)realis, counterfactual, subjunctive, etc., but does not encompass illocutionary-force or sentence-mood distinctions (declarative, imperative, etc.). As we saw above, where such contrasts are relevant to the morphological form of the verb, they were taken into account in the verb parameter.[10]

With these preliminaries in mind, we can now proceed to the parameter values that characterize the inflectional expression of TAM in the present sample:

- **Neutral**: There is no inflectional TAM marking in the language. Examples include the typically 'isolating' languages such as Chinese and Vietnamese.

- **Non-reduced**: TAM marking largely follows the same conventions as in independent clauses, i.e. the full or only a minimally reduced spectrum of TAM marking is possible. In language descriptions, this is typically reflected by statements that the complement clause "closely resemble[s a] normal finite clause, permitting almost the full range of verb inflections" (Evans (1995: 488) on Kayardild), or that the complement predicate "has no limitations on its TMA possibilities, nor is it dependent on the main predicate for its TMA organization." (van Gijn (2006: 310) on Yuracaré).

- **Particular**: This is a subtype of the non-reduced category above in which one of the TAM markers from the canonical main-clause paradigm is required to take on a particular value and is thus conventionally associated with the complementation pattern as a whole. For example, the Australian language Martuthunira has a specific kind of complement clause that can only be chosen under subject coreference and, crucially, needs to be marked for future tense:

(46) Martuthunira (Australian, Pama-Nyungan: Australia; Dench 1995: 256)

> *Ngayu   wiru   [yungku-ngu-layi   ngurnu-tharra-a   wirra-tharra-a].*
> 1SG.NOM   want   give-PASS-**FUT**   that.OBL-DU-ACC   boomerang-DU-ACC

---

[9] There is one exception to this, involving the category of mood: Some complementation patterns in the data must be marked by a specific modal particle (without this being an inflection). In such cases, I included the modal marking in my assessment of TAM (typically resulting in the value 'special' TAM marking (cf. below for this category)).

[10] Again, there is one exception to this general coding rule: It relates to direct-speech complements, which always constitute structurally independent clauses and hence follow independent TAM marking by definition. In those instances, the obligatory occurrence of, for example, imperative verbal morphology in jussive contexts, was exceptionally taken into account as 'differentially particular' TAM marking (again, see below for this category).

'I want to be given those two boomerangs.'

Similarly, perception complements in Huallaga Quechua need to be marked by Imperfective aspect (Weber 1989: 289). In one sense, such cases could alternatively be seen as restrictions of the TAM possibilities of main clauses, but what is criterial here is that the category itself can be expressed, and that it simply chooses a value from a canonical paradigm.

- **Differentially particular**: In this case, a TAM marker that is 'particular' by the criteria from above is limited to certain CTP classes, rather than applying to the complementation pattern as a whole.[11] In Rama, for example, an unmarked finite complement clause must obligatorily appear with Present-Tense inflection in perception and causative environments, regardless of the actual temporal interpretation of the complement (Grinevald 1990: 221f.). Similarly, jussive environments sometimes need imperative inflection (e.g. Hup or Jamsay), while desiderative contexts may trigger imperfective aspect, hypothetical mood and related values. In Skou, for instance, the predicate of a certain type of complement clause must exhibit a reduplication suffix to indicate irrealis mood, but only in desiderative environments:

(47) Skou (Skou, Western Skou: Indonesia; Donohue 2004: 443)

[*Hòe   nì=k-ang-kang*]   *nì=pung   li.*
 sago 1SG=1SG-eat-**REDUP**   1SG=liver   do
'I want to eat (some) sago.' ('liver do' = 'want')

Finally, differentially-particular TAM marking has also been found with regard to evidentiality (which is subsumed under TAM here). In Tariana, a quotative complementation strategy "must contain non-visual or inferred evidentiality" (Aikhenvald 2003: 554) when it occurs with propositional-attitude CTPs ('think, believe').

- **Special**: The complementation pattern exhibits a certain TAM marker that is specific to this construction or dependent clauses more generally and not part of the regular TAM-paradigm of basic declarative main clauses. *-cha/-na*-complements in Choctaw, for example, typically occur with a dependent aspect marker, the so-called 'L-grade', which is restricted to this construction and not used more regularly in the TAM paradigm:

(48) Choctaw (Muskogean: USA; Broadwell 2006: 275)

Si-ataklama-tok   [*Charles-at   iskáli'   habíina-na*].
1SG.II-bother-PST   Charles-NOM   money   receive.**LGR**-COMP:DS
'It bothered me that Charles received the money.'

- **Differentially special**: A TAM marker that is 'special' by the criteria from above is limited to certain CTP classes. For example, quotative-marked complements in Malayalam typically occur with a Debitive mood marker when they are used in jussive contexts, otherwise they are inflected regularly:

---

[11] Although I lack the relevant corpus data, I suspect that when the TAM of a complement is described as following the regular conventions of independent clauses ('non-reduced'), it will still be the case that, in language use, this typically boils down to differentially-particular marking since TAM is generally motivated in semantic terms: In other words, different CTPs will show at least a performance bias towards certain kinds of TAM marking. What my group 'differentially particular' captures is all those cases in which the specific marking on a certain CTP class is mentioned as noteworthy by the authors (i.e. explicitly pointed out as some sort of marked TAM potential).

(49) Malayalam (Dravidian: India; Asher and Kumari 1997: 46, 3)

    a.   [*Avan   satyasandhan aaɲə ennə*] *ɲaan   viʃvasikkunnu.*
        3SG.M   truthful.person be-PRS COMP   1SG   believe-PRS
        'I believe that he is honest.'

    b.   [*Ɲaan cellaɲam ennə*] *baabu paraɲɲu.*
        1SG   go-**DEB** COMP Babu say-PST
        'Babu said that I should go.'

- **Reduced**: The potential for the TAM inflections of independent clauses to occur is significantly reduced, normally in such a way that one or more of the otherwise obligatory TAM inflections cannot be expressed at all in the complement. Nominalized complement clauses in Basque, for instance, lose all TAM marking associated with finite clauses (Saltarelli 1988: 34):

(50) Basque (isolate: Spain, France; Hualde and Ortiz de Urbina 2003: 658)

    [*Haurrek   liburuak sarri irakurtze*]-*a-k*   *poztu egiten   gaitu.*
     children.ERG book.PL often read.NMLZ-DET-ERG cheer do.IPFV AUX
     'Children's often reading books makes us happy.'

Apart from such a wholesale reduction of all TAM categories, it is also possible for T, A and M to differ in their ability to be retained (cf. Comrie and Thompson 2007: 353, Cristofaro 2003: 127, Malchukov 2004: 80). In such cases, the coding proceeded along the cline T > A > M (i.e. priority was given to tense, then aspect, then mood). In Menya, for example, one complementation pattern in the database marks the same aspectual distinctions as independent clauses, but cannot be independently marked for tense; hence the whole construction is coded as TAM-reduced in comparison to main clauses (cf. Whitehead 2004: 177). By contrast, action nominalizations in Kwazá lose all kinds of mood and modal markers, but retain the possibility "to bear tense and aspect marking" (van der Voort 2004: 699); this combination clearly contrasts with the Basque-type of action nominal and since mood is outranked by tense and aspect, the pattern is coded as 'non-reduced'.

- **Inherent**: One or more of the TAM contrasts found in independent clauses is reduced to a relative-TAM contrast that is incorporated into the subordinator. For example, indirect-speech complements in Awa Pit are rendered by a converbal (called "participial") construction, and each of the two specific converbal morphemes to choose from is inherently specified for a particular "relative tense" (Curnow 1997: 260):

(51) Awa Pit (Barbacoan: Ecuador, Colombia; Curnow 1997: 261)

    a.   [*Miimal   puz-ta*]   *kizh-ti   María=na.*
        Chucunés   go:out-**PRFV:PTCP** say-TERM Maria=TOP
        'Maria said that (X) had gone out to Chucunés.'

    b.   [*Demetrio a-mtu*]   *kizh-ti   Carmen=na.*
        Demetrio come-**IPFV:PTCP** say-TERM Carmen=TOP
        'Carmen said that Demetrio is coming.'

The same goes for a number of nominalization constructions in the sample. For example, Muysken (1994: 2814) states in relation to Quechua that it "has maintained some tense distinctions in action nominals, which do not correspond to the past/present/future opposition of the main verb-tense paradigm, but to a distinction between realized

(past/present) and unrealized actions"; in Huallaga Quechua, this opposition is carried by the nominalizers *-na* and *-shqa*, respectively (Weber 1989: 288). In a recent typological systematization of nominalization processes, Gerner (2012: 808) refers to this phenomenon as 'suppletive nominalizers'.

These eight categories, then, reflect the principal ways in which TAM expression can be regulated in complement clauses. Let us now have a brief look at the cross-linguistic distribution of these TAM values in the data, specifically in relation to complementation systems again. Similarly to the verb forms above, it is instructive to examine how the TAM categories for specific complementation patterns combine into TAM values for entire languages.

To begin with, the 'neutral' value of TAM expression characterizes 7 of the sample languages and has a clear geographical bias towards the macro area of 'Southeast-Asia-and-Oceania': 5 of the 7 TAM-neutral languages are spoken in this area, viz. Chinese, Vietnamese, Karo Batak, Taba and Yakan; the three latter are Austronesian languages and hence also genetically related (though of partly very different families). The two remaining TAM-neutral languages are Hmong Njua, which counts as Eurasian in *WALS* but, being spoken in China, is clearly an East Asian language[12], and Abun, which is West Papuan by genealogy but spoken in Indonesia (and thus also not too distant from, for example, Taba). Therefore, we have a rather clear areal cluster here.

For the languages that have non-neutral TAM expression (N = 93), virtually all of the above TAM values and value combinations are attested in the data. Keeping the overall goals of the study in mind, there is no need to go into all these combinations here, but a few general observations should not go unmentioned. First, by far the most common TAM-pattern, found in 58 of the 93 languages (62%), is one in which relatively "balanced" (in Cristofaro's (2003) terms) TAM marking (i.e. non-reduced, particular or differentially particular expression) coexists with a severely reduced type (i.e. reduced or inherent marking; I am disregarding all '(differentially) special' cases here). This 'mixed' overall pattern is attested prominently in all macro-areas, and it no doubt reflects the association of the two TAM types in question with the morphological structure of the complement verb: balanced TAM expression is closely associated with independent verb forms, while reduced and inherent TAM often falls out from dependent verb forms that involve transcategorial operations, i.e. nominalizations, participles and converbs (Fisher exact test, p = 2.2e-16, odds ratio = 95.3).[13] The typologically interesting cases, therefore, are the ones that deviate from this expected pattern. These comprise, firstly, nominalizations that can still be marked for tense-aspect (e.g. West Greenlandic, Krongo, Kwazá, Mapudungun, Musqueam), illustrated for Mapudungun below:

---

[12] For the synchronic affinity between Hmong-Mien and Chinese languages more generally, cf., e.g., LaPolla (2001: 246).

[13] With regard to the other verb forms, 'bare stems' are TAM-deprived by definition, while our category 'other dependent verb form' shows a certain preference of balanced over reduced TAM inflection (9:6).

(52) Mapudungun (Araucanian: Chile; Smeets 2008: 214)

Kim-nie-n [*fey ñi ayü-**nie**-fiel*].
know-PROG-1SG.IND she 3POSS love-**PROG**-NMLZ
'I know that she loves him.'

Interestingly, the possibility of TAM-marking does not extend to the other transcategorial verb forms: All participles and converbs in my data are completely TAM-reduced.

The second deviation from expected patterns concerns verb forms that are not changed in their morphological status, but still involve a severe TAM reduction. This is found, for instance, in certain complements in Begak Ida'an, Lao, Supyire, Tariana, Tetun and Yuchi. The complements in question are often explicitly described as 'reduced clauses', 'control complements' or 'small clauses', and the typical pattern appears indeed to be such that there is some degree of "syntagmatic interweaving" (Lehmann 1988: 208) between the clauses in the form of argument sharing or other signs of morphosyntactic integration. Perception complements in Supyire, for instance, contain an independent verb form, but in contrast to other complementation patterns, they (i) cannot contain a subordinator, thus making matrix and subordinate clause immediately adjacent, (ii) need to code the matrix subject inside the complement but additionally also as the object of the perception verb ('interweaving'), (iii) are TAM-reduced, their inflectional possibilities being "severely restricted, allowing only the variation between perfective and imperfective aspect" (Carlson 1994: 424). An example of this construction is provided in (53) below:

(53) Supyire (Niger-Congo, Gur: Mali; Carlson 1994: 423)

Mìi a ù ɲyɛ̀ [*ú u ŋ-kɛ́ɛ́gé*].
I PRF him see he.COMP PROG INTR-go.IPFV
'I saw him going.'[14]

Returning to mixed TAM patterns more generally, another phenomenon that sticks out is the distribution of 'special' (and 'differentially special') TAM forms. In my data, these are never characteristic of languages as a whole: They occur in specific complementation constructions (N_{pat} = 17 in total) and these always co-exist with either regularly marked or reduced complements in a complementation system. The prevalent category that instantiates such special TAM marking is that of the 'subjunctive', i.e. understood here in a comparative sense as a modal inflection that specifically characterizes certain types of dependent clauses (cf. Noonan 2007: 61–65 for detailed discussion).[15] The distribution of subjunctives in my data is skewed quite heavily, showing a peak in African languages (Fongbe, Hausa, Kana, Koyra Chiini,

---

[14] Unfortunately, this specific example conceals the TAM-reduction somewhat (because the complement looks fully TA-marked), but the author's description is unambiguous and this is, of course, taken as criterial.

[15] Depending on the specific nature of the 'subjunctive' in a particular language, it may, of course, also be found outside of dependent clauses, often in non-canonical main clauses (imperatives, hortatives, etc., cf. Noonan 2007: 64f.), but its primary function is typically in dependent clauses.

Lango, Nkore-Kiga, Supyire) and Eurasia (Basque, Hungarian, Persian, Turkish)[16], but being rare everywhere else. It does occur occasionally in North America (e.g. the Slave Optative, the Purépecha Conditional) and South America (e.g. the Intentional marker on =*ti*-complements in Yuracaré with specific predicate classes, which is explicitly described to "function as a subjunctive" (van Gijn 2006: 319)); but it is absent from the major complementation patterns found in all other areas.[17]

As the final distributional observation, let me point out that apart from neutral and mixed-TAM systems in complementation, we also find languages with a predilection for either 'balanced' or 'reduced' TAM expression across all of their major complementation patterns. The balanced type (N_{lgs} = 22) is particularly characteristic of Northern American complementation systems (where it accounts for 7 of the 14 sample languages); in all other macro areas, it is considerably rarer. The reduced TAM type, by contrast, is infrequent across all macro areas (N_{lgs} = 8), appearing mainly in those languages whose complementation systems rely exclusively on nominalization and related processes (e.g. Dolakha Newar, Lavukaleve, Yimas, Kolyma Yukaghir).

This – admittedly very rough – distributional overview of TAM systems in complementation will suffice for the present purposes. At the end of this section, we will now turn to the more theoretical question of how the various TAM values from above are indicative of the decategorization processes that the present chapter intends to capture. It is well-known from the literature that TAM inflections serve to "particularize" (Croft 2007: 359) or "ground" (Langacker 1987: 126) a predicate so that it can express a specific instance of an event in the world. Conversely, a lack of TAM categories results in the "typification" (Lehmann 1988: 193) of the predicate again, which Lehmann lists as one dimension of desententialization more generally. The kinds of TAM marking distinguished above could now be seen as reflecting a continuum of decategorization again: The more contexts in which a given complementation pattern is used require the absence or special expression of TAM values, the more decategorized its status. This would suggest the following numerical coding schema of the data (Table 5):

---

[16] Note that German has a subjunctive in the above comparative sense, too, the so-called 'Konjunktiv I' in indirect speech, but this is a register-specific form that is virtually absent from indirect speech in the spoken language. Since this leaves the majority of indirect-speech complements TAM-marked regularly, German was not counted as one of the 'subjunctive' languages.

[17] Great care has to be taken when interpreting this finding. Specifically, it does not imply that modal inflections or modal markers more generally are less common in the other areas. On the contrary, as Palmer (2001) discusses in several chapters, many North American and especially Papuan languages have extensive modal marking (e.g. realis/irrealis contrasts), but the difference to the 'subjunctive' is that the latter "is found mainly in subordinate clauses" (Palmer 2001: 5), and it is this distributional property that is captured by 'special' TAM marking in my coding scheme above. Similarly, the Australian sample languages Gooniyandi and Mangarayi have been described elsewhere as having a subjunctive in their grammar (e.g. de Haan 2006: 33), but crucially, the descriptions of their complement clauses do not suggest any special modal marking that makes the complement deviate from an independent sentence (cf. McGregor 1990: 412f. and Merlan 1982: 9, respectively); therefore, their systems of modal morphology does not appear to fall under a strict definition of subjunctive.

Table 5. Numerical coding schema: Degree of decategorization of TAM inflections

1	Reduced or inherent TAM marking (= highly decategorized and typified)
0.5	Special TAM marking
0.25	Differentially special TAM marking
0	Non-reduced, particular or differentially particular TAM marking (= no decategorization or typification)

In either of the categories coded as '1' here, the predicate is characterized by the absence or severe reduction of TAM inflections and thus shows the highest degree of typification. At the other end, all cases coded as '0' involve TAM inflections from the canonical paradigm of independent clauses; despite preferences or conventional rules for particular TAM values, there is little reduction of the relevant inflectional category as such. 'Special marking' is located in between those two poles; certain types of modal marking, for example, occur on predicates that are 'finite' with regard to tense and aspect (e.g. the 'tensed *nú/ní*-complement' in Fongbe (Lefebvre and Brousseau 2002: 116)), but the presence of the modal morpheme contributes to marking the complement as a dependent entity. We may thus say that this situation shares features of both sentential and decategorized expression types (hence '0.5'). In differentially-special TAM marking, finally, the contexts in which decategorization takes place are limited to a certain number of CTPs, so in many instances, the complement will not be different from an independent sentence, which justifies the value '0.25'. As stated earlier, we will return to these values at the end of the chapter, when they are accumulated along with those of the other parameters.

4.3.2 Manner satellites

A type of predicate modification quite different from TAM inflections relates to the expression of modifiers of manner. These are normally expressed as phrasal satellites rather than inflectional operators. The basic contrast here, well-known in the literature on nominalization, is whether such manner satellites retain the form that they would normally take in independent clauses, i.e. that of an adverb phrase, or whether they succumb to nominalization and appear in the characteristic form of nominal modifiers, i.e. that of an adjective phrase. For example, we have already noted elsewhere that "complements in Yimas are clearly noun phrases" (Foley 1991: 394). Accordingly, they can take adjectival modification. (54a) below first introduces the general structure of such complements, showing that they exhibit a nominalizing morpheme followed by a noun-class marker that indicates the specific semantic type of complement. (54b) then shows the same structure with the CTP 'feel (like doing something)', and this time the complement predicate is accompanied by an attributive adjective:

(54) Yimas (Lower Sepik-Ramu: PNG; Foley 1991: 385, 395)

    a. [*Tpuk*   *am-tu-wampuŋ*]   *na-n-kacapal.*
       sago.SG eat-NMLZ-heart.CLF.SG   V.SG.OBJ-3SG.A-forget
       'He lost interest in eating sago.'

b. [*Kpa pacuk-t-wampuŋ*]   *ama-na-ti-n.*
  big copulate-NMLZ-heart.CLF.SG   1SG.S-DEF-feel-PRS
  'I am feeling very randy.' (lit. 'I feel like big copulating.')

There are literally only a handful of nominalizations in the present sample that work in the same way as Yimas.[18] Thus clear indications of adjectival coding were found in Georgian, Jarawara, Mapudungun and To'aba'ita. For many other nominalized complements (N_{pat} = 33), we have explicit evidence to the contrary. As a representative, we may cite Turkish nominalizations in *-DIK*, for which an overt statement is provided that "adverbials are not changed into adjectives" in the transcategorial shift (Kornfilt 1997: 57).

Interestingly, the literature on nominalization has also reported languages in which speakers vary between adjectival and adverbial modification (within an otherwise identical construction). Comrie and Thompson (2007: 374) list Polish and colloquial Egyptian Arabic as cases in point. Since this is also true of other varieties of Arabic (cf. Kremers 2003), it may well extend to Gulf Arabic in our sample (although I have no conclusive evidence here). Similarly, Malayalam nominalizations in *-al* and its related forms can have adverbial manner satellites, but Andronov (1996) reports that the construction with *-al* is becoming increasingly noun-like, thus losing some of its verbal properties. Accordingly, in Jayaseelan (1999), *-al*-constructions are already considered NP with adjectival modifiers. In view of the indications of variability, Arabic and Malayalam will be coded as having a 'mixed' pattern of manner modification.

This brings us to the numerical coding of the variable, which is very straightforward in this case (Table 6):

Table 6. Numerical coding schema: Form of manner satellites

1	adjectival (= full recategorization)
0.5	mixed (= optional recategorization)
0	adverbial (= no decategorization)

It has to be conceded, however, that there are quite a few cases in the database for which no reliable data were available on the form of manner satellites in nominalizations. I decided to code these instances conservatively, subsuming them under the 'default' value '0': Since what we ultimately want to capture is the degree of recategorization, the construction in question simply should not be treated as contributing to it if there is no evidence thereof.

4.4 Actant modification

In this section, we are turning to what Bossong (1979: 40) calls 'actant modification', i.e. to the changes in form that the arguments of a complement clause may undergo in the processes of de-/re-categorization. Phenomena like possessive coding of

---

[18] This is not particularly surprising, if we recall that adjectival coding was one of the criteria to distinguish complement clauses proper from complementation strategies, and that the latter entered the present database only under special circumstances (cf. §3.2).

complement subjects and objects have been well-studied in the typological literature on nominalization (e.g. Koptjevskaja-Tamm 1993), while other syntactic processes such as raising and control have an extensive research tradition on their own, both inside and outside typology. Even cross-linguistic survey articles on complementation sometimes devote considerable space to these phenomena, as Noonan (2007) illustrates. However, following the multivariate agenda of the present work, the topic can still be approached from a novel perspective. Specifically, we can try to typologize the argument-structural arrangements in complement clauses by dividing them into a number of logically independent parameters again, and by examining the cross-linguistic combinations of these parameters. As will be seen, this is an enormously complex enterprise, and the present section cannot by any means fill this gap exhaustively (without digressing too much from the overall agenda of the current chapter). However, it will try to make a first step into this direction. The general idea is that Noonan's survey work on the topic can be elevated from a purely qualitative overview to an empirical study, and that certain theoretically relevant points for the typology of complementation can be recognized as we go along.

The section is organized as follows. In §4.4.1, I will focus exclusively on the subject of the complement clause and outline the various processes that may affect its coding properties in different kinds of complementation environments (e.g. same-subject or different-subject constellations, subject versus object clauses, pronominal indexes versus lexical NPs, etc.). This will result in a set of parameters that, taken together, characterize the expression of the internal subject for each complementation pattern in the database. In §4.4.2, I will look at the coding of the internal direct object argument. This will need a much less detailed exposition than the subject parameters, but it is still significant for capturing differences between individual complementation patterns in the data. As we proceed, all parameters will also receive a numerical assessment of the degree of re-/de-categorization they entail. In §4.4.3, finally, I will use exploratory statistical methods to investigate combinations of argument-structural properties, so as to identify cross-linguistically recurrent ways of subject and object expression in complementation.

4.4.1 The coding of the complement subject

4.4.1.1 Same-subject contexts

Given that complements express a relatively intimate semantic relationship between two events, one being construed as an argument of the other, the sharing of participants between the events in question is not at all unusual. Consequently, a number of the complement-taking environments distinguished for the present study often involve a certain type of coreferentiality between the internal arguments of matrix and complement clause. The present section examines the expression of the complement subject argument where this is coreferential with the subject argument of the matrix clause. It goes without saying that this scenario only applies to object clauses: Since subject clauses, as a whole, constitute the main-clause subject, they cannot be coreferential with a subject participant in the matrix. In object clauses, by contrast, the issue is highly relevant. For some of our environments, such as the

'phasal-P' group, this type of coreference is a logical necessity (*I began _their_ playing the symphony_). For others, such as desiderative complements, it has been shown to be the dominant type in performance, outranking different-subject contexts significantly in its occurrence frequency in several languages (cf. Schmidtke-Bode 2012, Haspelmath to appear). Perception verbs, by contrast, are not entirely impossible with coreferential subjects (_I heard myself say stupid things_), but normally involve a different-subject constellation (_I saw him go to the library_). Overall, most of the CTP classes distinguished in this dissertation do not logically "predetermine" (Cristofaro 2003: 116) the identity of the subjects and are hence open to both same-subject and a different-subject constellations. Therefore, the ratio of same to different subjects for these predicate classes will vary in the corpora of individual languages, often depending on sociocultural and discourse conventions. In Supyire, for instance, same-subject constellations with knowledge, prepositional-attitude and quotative complements are noticeably frequent in performance simply because "people like to talk about themselves" (Carlson 1994: 444).

The major issue for a typology of complementation is, of course, how the complement subject is expressed under coreference with the matrix subject, i.e. how this affects the form of the complementation pattern in question. Of the 228 complements in my data, 31 (N_{pat} = 13.6%) are such that I did not find evidence for them being used in same-subject contexts, or we know that they are not normally used in such environments to begin with. It is clear, for example, that a medial clause marked by a different-subject switch-reference morpheme, such as the perception complement in Menya, cannot be used in same-subject contexts. For the remaining data points (N_{pat} = 197), the expression of coreferential subjects draws on a number of different techniques, each of which is outlined in the following.

One option, characteristic of one third of all data points (N_{pat} = 82/228 = 36%), is to express the coreferential subject **canonically**, i.e. in the same way that it would take in independent clauses. If the primary expression format for subjects in a language is indexation (as in so-called 'pro-drop' languages), this results in both the matrix verb and the complement verb being indexed according to the regular rules for independent clauses.[19] This is shown in (55) below.

(55) Tamashek (Afro-Asiatic, Berber: Mali; Heath 2005: 695)

  `I-nna   [i-tátt   dihá].
  3SG.M.SBJ-say.PRFV.P   **3SG.M.SBJ**-eat.IPFV.P   here
  'He_i says that he_i eats here.'

On the other hand, (56) shows an example of a language that relies on free pronouns rather than indexes to express the subject, and where these are overt in both the matrix and the complement:

---

[19] Determining the 'primary type of subject expression' largely followed and greatly profited from Dryer's (2011a) chapter 101 in _WALS_. Where languages had both subject indexation and free subject pronouns, an individual decision was made that is tailored to the complementation system. For example, indexation is generally left overt on complement clauses in Tetun, but the individual complementation patterns differ as to whether they allow free subject pronouns in addition to that. Since the latter criterion thus distinguishes different data points from one another, the expression of free pronouns was taken as criterial.

(56)   Mayogo (Niger-Congo, Adamawa-Ubangian: DR of Congo; Sawka 2001: 160)

Ma nga-kʉnda [*me   ma nʉ bhʉ gʉdhʉ*].
1SG PROG-like   COMP **1SG** go   to   town
'I want to go to town.'

Similarly, in one complementation pattern in Abun (West Papuan: Indonesia), coreferential subjects can be deleted, but "this type of ellipsis is not common, as it is more usual that the subject is repeated." (Berry and Berry 1999: 179) In such cases, the preferred choice was taken as criterial for the coding process.

Another option is, of course, to conventionalize the omission of the coreferential subject as a syntactic constraint. This is what is widely known as '**control**' or '**equi-deletion**' (cf. Rosenbaum 1967, Landau 2000 for two influential formal works on the topic). In control constructions, the complement subject cannot be expressed overtly, so that the complementation pattern acquires a structural characteristic that makes it different from independent clauses.[20] In my sample, about 32% of the data points (N_{pat} = 72/228) involve control processes in same-subject constellations. This can again relate to both free pronouns and indexes on the complement verb, as shown in (57) and (58) below:

(57)   control of subject indexation in Tukang Besi (Austronesian, Western-Malayo Polynesian, Sulawesi: Indonesia; Donohue 1999b: 389)

Ku-hada [*te   wila'-a   i   'one*].
1SG-want   CORE go-NMLZ   OBL beach
'I want to go to the beach.'

(58)   control of subject pronouns in Warao (isolate: Venezuela; Romero-Figeroa 1997: 13)

Oko [*hi   kaika nao-kitane*]   *obono-bu-te.*
we   2SG.OBJ with come-CONV   want-ITER-NPST
'We really want to come with you.'

Control phenomena have received ample attention in the literature and their qualitative dimensions need not be discussed any further here (cf. also Stiebels 2007 for a recent synopsis). From a quantitative point of view, however, my data allow for at least a rough estimate as to how and how widely control processes are distributed in the world's languages.[21] Noonan (2007: 78) maintains that "equi-deletion is a common process", and this claim can now be substantiated empirically: In all six macro areas, the *majority* of sample languages shows evidence of coreferential-subject control, although the exact figures differ considerably. Thus while *all* of the African languages

---

[20] In this definition of equi-deletion processes, I follow Noonan (2007: 78) and Andrews (2007a: 171). It contrasts with the approach to control taken in a recent cross-linguistic study (Stiebels 2007), where a broader, semantically-based definition of control relationships is assumed that does not necessarily result in deletion processes. Furthermore, it also contrasts with Falk's (2006: 135) concept of control, which comprises both control in the present sense and raising processes, the latter of which will be dealt with separately in the present study.

[21] The estimate is called "rough" here because it has to be kept in mind again that my database of complementation patterns records the major complementation techniques of each language, and only those for which sufficiently detailed information was available. I have, additionally, taken notes on constructions that were not (or could not be) taken on board, and these were integrated into the calculation as far as possible. But there is, of course, always a residual risk of having missed a certain number of relevant constructions in individual languages.

in the sample exhibit control processes, the figures are lower for Eurasian and Southeast-Asian languages (slightly over 70% of the languages in both areas), and yet lower for the Americas and Australia-New Guinea (between 57% and 59% of the sample languages from these areas). I refrain from producing more detailed numbers and significance tests because of the caveats mentioned in footnote 21; clearly, more detailed work is needed in the future to provide a more reliable picture of this issue.

Returning to the omission of coreferential subjects, it is vital to distinguish control constructions from situations in which the complement subject is not expressed overtly but without this being an obligatory syntactic property of the complementation pattern in question. As is well-known, languages differ dramatically in what Bickel (2003) calls their "referential density" in discourse, and it holds in particular that many languages freely omit subject arguments in independent clauses. As subjects tend to encode discourse-old information, i.e. easily recoverable referents (cf. Givón 1979: 26f. and many others), economical considerations on the part of the speaker suggest leaving such topical subjects implicit. If this already applies to regular independent clauses, then the omission of subjects in same-subject complement clauses would only be expected for these languages. However, the crucial point is that the resulting structure of the complement does not differ in principle from an independent clause; in contrast to control processes, then, no decategorization of the complement takes place. This is described, for example, for so-called 'paratactic complement clauses' in Semelai (Austro-Asiatic, Mon-Khmer, Aslian: Malaysia), where "the optional elision of a core role from the complement clause is motivated by anaphoric discourse elision" (Kruspe 2004: 348) and not driven by syntactic control processes.

In my sample, the most relevant instances of this pattern come from languages that do not have indexation systems but still freely omit their subject pronouns in discourse. Thus languages like Burmese, Epena Pedee, Hup, Imonda, Japanese, Kayardild, Modern Khwe, Korean, Lao, Lezgian, Malayalam, Mandarin Chinese or Tümpisa Shoshone often feature the absence of a coreferential subject in some of their complementation patterns, but this is inconspicuous as compared to independent clauses in the same language. Sometimes, however, it is possible that even in such languages, a certain type of complement *cannot* express coreferential subjects and thus exhibits control (e.g. 'non-finite' constructions Lao, Lezgian or Malayalam, to name but three). A careful analysis of language-level and construction-specific properties was thus necessary in order to arrive at an appropriate characterization of the coreference parameter for each complementation pattern. Overall, the number of **freely omitted subjects** in my data amounts to $N_{pat} = 23/228 = 10\%$. If we add the genuine control constructions from above ($N_{pat} = 72$), then the total number of complementation patterns with implicit coreferential subjects (as a grammatical rule or a performance preference) rises to $N_{pat} = 95/228 \approx 42\%$.[22]

---

[22] Intuitively, one would perhaps have expected this figure to be higher. It has to be kept in mind, though, that this is not a figure for individual CTP classes, but for complementation patterns as a whole. It is, of course, true that same-subject desiderative environments will have a higher percentage of implicit-subject complements than propositional-attitude or quotative CTPs. But this is an entirely different perspective on the data, which will be pursued properly in Chapter 6 of the dissertation.

Apart from overt, controlled and freely omitted subjects, another possibility is for the coreferential subject to take on a **special** form as opposed to independent clauses. A well-known case in point is the possessive coding of subjects in nominalizations. In general, the nominalization patterns in my data overwhelmingly prefer equi-deletion (N_{pat} = 45/65 = 69%) to the possessive coding of coreferential subjects (N_{pat} = 6/65 = 9.3%).[23] The latter is found, for example, in Musqueam, Huallaga Quechua, Turkish and Wari'. (59) provides a pertinent example:

(59) Musqueam (Salishan, Central Salish: Canada; Suttles 2004: 102)

　　　Sí·səy cən [*k̓ʷə　nə-s-ném̓*].

　　　fear　I　　ART　**my**-NMLZ-go

　　　'I'm afraid to go.' (lit. 'I am afraid of my [hypothetical] going.')

Another type of special subject coding is found where languages employ a switch-reference system in complementation, so that the canonical subject indexes are replaced by same-subject markers.[24] In a way, this is the argument-structural counterpart to 'relative tense' in the TAM domain because the subject is not expressed in absolute terms but indicated relatively to the matrix subject. This happens in the converbal clauses in Matsés that we encountered earlier; thus whereas in examples (38a–b) from above, the converbal 'when'-suffix is inherently specified for different subjects, other converbal endings (e.g. 'while' for phasal verbs) are inherently marked for same subjects (Fleck 2006: 237). Similarly, in one type of complement in Choctaw, a switch-reference marker is normally added to a fully inflected verb, but under subject coreferentiality, the regular subject index is frequently dispensed with, so that the switch-reference marker is left as a relative signal of subject expression (cf. Davies 1981: 43 and Broadwell 2006: 282):

(60) Choctaw (Muskogean: USA; Broadwell 2006: 269)

　　　John-at　anokfilli-h [*pisachokma-ka-t*].

　　　John-NOM　think-TNS　goodlooking-COMP-**SS**

　　　'John$_i$ thinks that he$_i$ is good-looking.'

A much more prominent type of special subject marking is that of logophoricity and closely related phenomena. Logophoricity "involves special third-person pronouns occurring in dependent clauses only and expressing coreferentiality with the subject of the main clause." (Creissels et al. 2008: 144; cf. also Culy 1994 or Güldemann and von Roncador 2002 on this phenomenon). This is illustrated for Jamsay below, whose indirect-speech complement uses a distinct pronominal form for third-person coreferential subjects:

---

[23] The remaining 14 nominalizations fall into the other categories discussed above or have mixed values for the expression of coreferential subjects.

[24] Where, by contrast, switch-reference marking does not replace the subject indexes but occurs in addition to them, switch-reference marking does *not* entail 'special' subject expression. In such cases, the form of the indexes is taken as criterial.

(61)   Jamsay (Niger-Congo, Dogon: Mali; Heath 2008: 587)

    [ènɛ́   wó   wɔ̀ː-ø]   wà.

    LOG.SBJ   3SG.OBJ kill.IPFV-3SG.SBJ   say

    'He$_i$ said (to her$_j$) that he$_i$ would kill her$_j$.'

There are several variations on this pattern: In Lango, logophoricity is a matter of indexation rather than free pronouns. In Krongo and Supyire, a set of emphatic personal pronouns is recruited for logophoric purposes. In Kana, logophoricity (in the form of a special affix) extends beyond strictly quotative constructions and beyond the third-person context; in other words, all environments in which the complementation pattern appears ('want', 'know', etc.) trigger logophoric marking if the subjects are coreferential:

(62)   Kana (Niger-Congo, Benue-Congo, Cross-River: Nigeria; Ikoro 1996: 284)

     M̀-dā̠ā̠   [kɔ̄ɔ̀   ṁm̀-ɛ̀   nɛ̄ɛ̀-péé].

    I-hear.FACT   CONN   I.COP.PRS-**LOG** person.ASSOC-goat

    'I heard that I am an idiot.'

Logophoricity has been identified as a robust areal feature within the African continent (cf. Heine 2009a: 5), where it is common "in an area ranging from Senegal in the west to Ethiopia in the east, and cutting across genetic boundaries" (Creissels et al. 2008: 144; cf. also Güldemann 2003 for specific data). However, it can occasionally also be identified in other macro areas, such as Lao in Southeast Asia (Tai-Kadai, Kam-Tai: Laos, Thailand; Enfield 2007: 430), Yagua in South America (Peba-Yaguan: Peru; Payne and Payne 1990: 335) and Wappo in North America (Wappo-Yukian: USA). The latter has a special third-person pronoun for same-subject contexts, used across the board in all subordinate functions that the construction in question can cover (e.g. also in adverbial clauses, cf. Thompson et al. 2006: 109). In sample languages from Eurasia, logophoricity comes in the form of reflexive pronouns. This is obligatory in third-person contexts of indirect speech in Lezgian (Haspelmath 1993: 354), and optional for all complement clauses in Korean (cf. Gamerschlag 2007, Park 1995). In Evenki, finally, both complementation constructions in the data occur with possessive subject marking, but this needs to turn into possessive-reflexive marking under subject coreferentiality:

(63)   Evenki (Altaic, Tungusic: Russia; Nedjalkov 1997: 25)

    Nungan   sa:-re-n   [eme-d'enge-vi].

    he   know-NFUT-3SG   come-PTCP-POSS.**REFL**

    'He$_i$ knows that he$_i$ will (be able to) come.'

Taken together, the various forms of 'special' subject marking in same-subject environments account for $N_{pat} = 30 = 13\%$ of the data. This figure includes all cases like Korean, where special marking is an option that alternates with other types of subject expression in the same complementation pattern. This brings us, finally, to 'mixed' strategies of same-subject coding, which are basically combinations of the above possibilities. For example, in the South American language Trumai, indexation is available for third-person Absolutive arguments, while all other pronominal forms are rendered by free personal pronouns. In so-called 'Dative complement clauses', the

Absolutive index must be replaced by a possessive index (creating 'special' subject marking), but the personal pronouns of all other contexts remain canonical. Therefore, when it comes to assessing the expression of coreferential subjects, we end up with a 'mixed' pattern, as illustrated below:

(64) Trumai (isolate: Brazil; Guirardello 1999: 381, 379)

a. *Kiki yi iʈï ka_in* [*fakdits-ea*]=*tl.*
   man *yi* fear FOC/TNS die-**3POSS**=DAT
   'The man₍ᵢ₎ fears he₍ᵢ₎ will die.'

b. *Ha pudits ka_in* [*ha otl yi*]=*ki.*
   1 like FOC/TNS **1** sleep *yi*=DAT
   'I like sleeping.'

In all the preceding calculations, the individual values of mixed patterns were included in the figures for 'canonical', 'deletion', 'free omission' and 'special', respectively.

We are now in position to evaluate the different types of same-subject expression from the perspective of de-/re-categorization processes in complementation. I have already pointed out that both the canonical type of subject expression (i.e. regular indexation or pronouns) and the cases of freely omitted subjects do not constitute an aberration from independent clauses. They contrast in this respect with obligatory subject deletion (control processes) and those types of 'special' subject expression that mark the complement clause as a dependent entity (e.g. possessive encoding, and switch-reference marking). In keeping with our earlier parameters, we can thus distinguish two opposite ends of the decategorization continuum.[25] The other expression types need to be considered individually again. Constructions with logophoric marking in the prototypical sense are decategorized for third-person coreferential subjects since these receive a treatment that is specific to dependent clauses. Non-third-person contexts, by contrast, usually follow canonical rules. Therefore, classic logophoricity results in partial decategorization. This also holds for the 'mixed' patterns of subject expression, as long as one of the contributing values is of the unmarked type (i.e. canonical, freely omitted subjects or logophoric). Where mixed constructions involve only decategorizing operations (e.g. control of indexation and possessive coding of free pronouns), they were classified as decategorized overall. This policy results in the following numerical coding scheme for the present parameter:

---

[25] Note that the *re*-categorization induced by possessive markers will not be ranked more highly than other types of deranked subject expression (just like nominalized verb forms did not outrank participles or converbs in §4.2.7 above). Such an internal ranking would be particularly unfortunate here since we saw that the primary expression type for coreferential subjects in nominalizations is clearly 'deletion', and these cases should not count as less de/re-categorized than the few cases with overt possessive markers.

Table 7. Numerical coding schema: Expression of coreferential subjects[26]

1	Obligatory deletion (control), possessive and relative (switch-reference) coding; mixed patterns with combinations of these values (= highly decategorizing and partially also recategorizing)
0.5	Logophoricity and mixed patterns with partially independent expression types
0	Canonical expression and free subject omission (= no de-/re-categorization)

4.4.1.2 Different-subject contexts

The expression of complement subjects that are not coreferential with the matrix subject basically also relies on the techniques introduced so far (except for logophoricity). Nevertheless, the overall picture is more complex because (i) the scope is expanded to incorporate subject clauses, (ii) new techniques need to be recognized (e.g. raising), (iii) a multivariate coding procedure is required due to the coexistence of different techniques within a single complementation pattern. Furthermore, different-subject contexts are quite heterogeneous, so that two principal types of constellation need to be distinguished:

(a) The complement subject is referentially identical to a non-subject argument in the matrix clause (e.g. the direct object).

(b) The complement subject codes a referent that has no role to play in the matrix situation.

Although complementation structures do not normally fall neatly into these two categories (often being applicable to both contexts), it will still be useful to organize the presentation of different-subject contexts around this distinction. Therefore, I will first describe the various phenomena that can be encountered in constellations (a) and (b), before I present a coding scheme that breaks the notion of different subjects down into several distinct parameters.

(a) *Referential identity of the complement subject and a matrix non-subject argument*

Perhaps the first contexts that come to mind for this scenario are manipulative ones: In both jussive and causative situations, a non-subject participant of the matrix clause is ordered to do or coerced into doing something, so that s/he is also normally the subject participant in the resulting event. Taking up our previous distinctions again, this cross-clausal argument sharing may invite **equi-deletion** of the complement subject. In such cases, the matrix object functions as the 'controller' of the complement subject, which is why this type of deletion process has also been called '**object control**' in the literature. The 'controllee' is then obligatorily left implicit, as indicated by the absence of a subject index or free personal pronoun. In Gulf Arabic, for example, nominalizations are normally indexed overtly by possessive markers, but the index is

---

[26] Complementation patterns that do not allow or are not attested with coreferential subjects will not be coded numerically, but will instead be left in their categorical form. In this way, they cannot increase the overall index of decategorization or nominality for a given complementation pattern.

removed in jussive contexts (Holes 1990: 21). In German *zu*-Infinitives, indexation is generally eliminated, and no subject pronouns are allowed inside the complement. In jussive contexts, the Infinitival subject is controlled by a Dative object in the matrix clause:

(65) Object control in German (Indo-European, Germanic: Germany, Austria, Switzerland)

Ich   befahl   ihm   [zu   geh-en].
1SG.NOM   order.3SG.PST 3SG.M.DAT to   go-INF
'I ordered him to leave.'

Of course, it is also possible to leave the complement subject in **canonical** form. This can be seen with regard to subject indexation in Taba. In the following example, the patient of the ordering event is coded as an applicative object in the matrix and then expressed again as a canonical index on the complement verb:

(66) Taba (Austronesian, Eastern-Malayo Polynesian, South Halmahera-West New Guinea: Indonesia; Bowden 1997: 444)

N=sul-ak   wang=si   [de   l=mul   ak-le].
3SG=order-APPL   child=PL   COMP **3PL**=return   ALL-land
'He told the children to go home.'

Special types of subject expression in manipulative contexts include the ones from above (chiefly possessive coding), but also indexation paradigms that are specific to dependent clauses and thus signal their subordinate status. (67) provides an example from Musqueam:

(67) Musqueam (Salishan, Central Salish: Canada; Suttles 2004: 95)

Cəs-ét-ø   cən ceʔ   [wə-nə́ṁ-əs].
tell-TR-3P   I   FUT   COMP-go-**3SUB**
'I'll tell him to go.'

The type of argument sharing that is characteristic of manipulative CTPs with object clauses also extends to subject clauses, notably those with evaluative and deontic CTPs. In the following examples, we first return to the German *zu*-Infinitive from above to illustrate equi-deletion (ex. (68)), before we look at a case of canonical subject marking from Gulf Arabic (ex. (69)):

(68) German (Indo-European, Germanic: Germany, Austria, Switzerland)

Es   ist   wichtig für mich, [dies-es   Buch   zu schreib-en].
it.NOM be.3SG.PRS important **for** 1SG.ACC this-N.SG.ACC book.N.SG.ACC to write-INF
'It is important for me write this book.'

(69) Gulf Arabic (Afro-Asiatic, Semitic: Kuwait etc.; Holes 1990: 111)

ṣaʕb ʕaleek   [ʔinnak   tinaggil   han-nagaayil].
hard   on.**2SG** COMP.**2SG** **2SG.M**.transplant these-seedlings
'It is hard for you to transplant these seedlings.'

Without a controller in the matrix clause, the absence of subject marking inside the complement clause often enforces a generic interpretation[27]:

(70)　Tümpisa Shoshone (Uto-Aztecan, Numic: USA; Dayley 1989: 375)

　　　[*Na-maapüatu-nna*]　　*tsawinnuh.*
　　　 PASS-help-INF　　　　 be.good
　　　'To be helped is good.'

Noonan (2007: 78) argues that such constructions do not involve control "since conditions for coreference have not been met, i.e. there is no matrix argument which the subject can be identical to." On Stiebels' (2007) approach, they typically involve "implicit generic control" since a controller could potentially be supplied. In order to sidestep this terminological controversy, I will not use the term control but simply stick with '**generic-implicit**' subjects. In fact, this term will be generalized to all kinds of implicit subjects in subject clauses since for many of the relevant examples in my data, only the generic variant is attested and it is unclear whether an overt controller could be added to the matrix clause.

　　　Finally, it is also possible to find 'special' marking in subject clauses of this type. Interestingly, this does not only apply to overt subjects (whose form might, for example, change into a possessive index again), but also to subjects in generic contexts. Thus in West Greenlandic, a so-called "impersonal 4th person" affix (Fortescue 1984: 39) is used when no overt controller is given; this affix has reflexive or logophoric functions in object clauses. Its use in subject clauses is illustrated in (71):

(71)　West Greenlandic (Eskimo-Aleut: Greenland; Fortescue 1984: 39)

　　　[*Mianirsur-luni*]　　　　 *ajunngin-niru-ssa-aq.*
　　　 be.careful-**4SG**.CTP　　 be.good-more-FUT-3SG.IND
　　　'It would be best to be careful.'

In Tariana, too, a special impersonal marker can be found, but this time it is triggered by an overt controller in the matrix clause:

(72)　Tariana (Arawakan: Brazil; Aikhenvald 2006: 189)

　　　[*Du-haniri-ne*　　　 *pa-sape-hyu*]　　　　 *manhina-ma-na*　　　　 *nu-na.*
　　　 3SG.F-father-INS　 **IMPS**-speak-PURP.NVIS　 be.hard-EXCES-REM:PST.VIS　1SG-OBJ
　　　'It was hard for me to speak with her father.'

(b)　*Disjunctive reference of the complement subject and the matrix arguments*

In the second type of participant constellation in different-subject environments, the complement subject codes a referent that has no semantic role to play in the matrix situation. This scenario calls for the complement subject to be overtly mentioned, either in canonical or one of the special forms introduced previously. This is illustrated

---

[27] This is, of course, particularly characteristic of nominalizations and other dependent verb forms. According to Comrie and Thompson (2007: 368), "apparently in all languages with action nominalizations, it is possible to leave the subject unexpressed, the nominalization then referring to an abstract type of activity or state". In Noonan (2007: 118), these structures are called 'activity nominalizations', while the ones with explicit subjects are termed 'nominalized propositions'.

for two subject clauses below, where one marks the subject as in independent clauses (73) and the other adapts it to the nominalized character of the complement (74):

(73) Ndyuka (Creole: French Guiana, Suriname; Huttar and Huttar 1994: 75)

A	*bun*	[*taki*	*u*	*go*	*a*	*Soolan*].
3SG	good	COMP	**1/2PL**	go	LOC	St. Laurent

'It's good that we went to St. Laurent.'

(74) Turkish (Altaic, Turkic: Turkey; Johanson 2011: 7)

[*Ayşe'nin*	*gel-diğ-i*]	*mümkün.*
Ayşe.**GEN**	come-ACT.NMLZ-3SG.**POSS**	possible

'It is possible that Ayşe comes/came/has come.'

The possessive type of subject coding seen in the Turkish example is, of course, a common solution for nominalized complements in different-subject contexts. It needs to be pointed out, however, that quite a few nominalizations in the data belong to what Koptjevskaja-Tamm (1993) calls the 'sentential' type of action nominal: They display morphological nominalization of the predicate, but without concomitant actant modification. This applies to the nominalized complement clauses in Basque, Hup, Kwazá, Lavukaleve, Lezgian and Tariana. An example from Lezgian was provided earlier (cf. (28) in §2.3.2), and one from Hup is given here:

(75) Hup (Vaupés-Japurá: Brazil, Colombia; Epps 2008: 850)

[*Tɨh*	*tɔʔóh-nɨ́h*]	*ʔãh*	*tuk-nɨ́h=hɔ̃.*
3SG	run-**NMLZ**	1SG	want-NEG=NVIS

'I don't want him to run (away).'

An important observation in this connection is that different techniques for subject coding may go their own ways within the same complementation pattern. In the context of nominalizations, Malchukov (2004: 37) refers to this phenomenon as "operator-satellite mismatches". Thus some nominalized complements may delete their subject indexes (i.e. inflectional operators) while the corresponding satellite expressions (i.e. the conominal argument NPs) may be retained in canonical form. This was illustrated in (50) above from Basque, and we also find it in (certain) nominalizations in Amele, Barasano, West Greenlandic, Korafe, Dolakha Newar, Huallaga Quechua and Wolaytta, as well as in the converbal construction of Awa Pit from (51b) above. Operator-satellite mismatches can also take other shapes. For example, in nominalizations in Gulf Arabic, pronominal subjects are coded as possessive indexes, while lexical subject NPs retain their canonical form (Holes 1990: 22). The converb in Matsés encountered earlier (cf. (38)) shows special (switch-reference) indexation but again a full-fledged subject NP. More generally, then, the occurrence of such cases suggests breaking the different-subject parameter down into several individual variables, which is exactly that we will do at the very end of this section.

Apart from the possessive and switch-reference strategies, there is another type of non-canonical subject marking that we have not come upon yet. It relates to the cases in which the complement subject "trespasses" its clause boundary (Hawkins 1986: 94) to contract a syntactic relationship with the matrix predicate, even though it is not a

thematic argument of this predicate. This phenomenon is most widely known as **raising** or, alternatively, as "matrix coding" (Van Valin and LaPolla 1997: 561). Similarly to control processes above, we will have to devote some space to a discussion of raising at this point, for several reasons. First, raising is an intricate phenomenon that needs to be distinguished from functionally equivalent constructions which are syntactically different (notably prolepsis). Secondly, raising applies to both subject and object clauses, each of which deserves individual attention. Third, despite the fact that Noonan's (2007) overview of complementation discusses raising processes quite extensively, it does so without drawing on a principled database (i.e. without quantification) and without noting the implications of raising (and related) processes for the typology of complementation systems. It is only after examining these issues carefully that we can settle on the coding of the different-subject parameter for the relevant complementation patterns. Therefore, I 'interrupt' the different types of subject marking by an excursus into raising.

Excursus: Raising

We begin the discussion by looking at 'subject-to-object' (SO-) raising (Postal 1974) in object complements.[28] As the term implies, the subject of the complement clause ends up syntactically as an object argument of the matrix predicate. This is typically indicated by case marking appropriate to a matrix object (as in (76) below) or explicit object indexation on the CTP (77):

(76)   Krongo (Nilo-Saharan, Kadugli: Sudan; Reh 1985: 337)

N-átàasà   *à²àŋ ù²ùŋ*   [*k-áaláaná   à²àŋ kí-nìinò   mó-dì*].
1/2SG-want.IPFV I   **you.ACC**   LOC-teach.INF me   LOC-language GEN-home
'I want you to teach me Krongo.'

(77)   Motuna (East Bougainville: PNG; Onishi 1994: 488)

[*Hoo   lootu   poko-no-ita   jii   hia   roki=manni topo*
 ART.M prayer 3SG.POSS-LINK-CLF.side and   thing really   well

kuuk-arei-ko]   *haa=haa-jee-m-a-i* ...
know-NMLZ-EMPH   REDUP=want-**APPL-1P**-3PCL.A-CONT.SS
'They always wanted me to know things related to Christianity really well ...'

As can be seen, raising leads to a non-iconic mapping between syntactic and semantic roles since the raised argument does not have a semantic role in the matrix situation. Consequently, a key feature of SO-raising is that the matrix proposition is not entailed by the proposition of the entire sentence (*I want you to go* ⊭ *I want you*).

   In the literature, SO-raising has sometimes also been applied to verbs of perception (e.g. Eisenberg (2006: 368) on German (e.g. *Ich sah sie gehen*), Genetti (2006: 142) on Dolakha Newar). In contrast to the above cases, however, perception verbs and related

---

[28] Note that the entire exposition will focus on the morphosyntactic aspects of raising process; their motivation in terms of semantic and discourse-pragmatic factors is amply discussed elsewhere (e.g. Langacker 1995, Givón 2001, Serdobolskaya 2009) and will not concern us here.

predicates take the complement subject as their object argument syntactically *and semantically*, with the matrix proposition being entailed (*I saw you leave* ⊨ *I saw you*). Other authors have argued, therefore, that perception complements of this kind exhibit equi-deletion rather than raising (e.g. Arkadiev (2012: 313ff.) on Participial perception complements in Lithuanian, König and Gast (2012: 235) on German Infinitives after perception verbs). In the present study, evidence for 'matrix coding' in the context of perception verbs will be treated as a "raising effect" in Serdobolskaya's (2009) sense: it will be coded as a kind of non-canonical subject marking (so as not to lose the special pattern in the data), but I noted whether this effect only applies to perception verbs only or whether it is extended to CTPs for which a genuine raising analysis must be assumed.

The 'mild' raising effect observed in perception complements contrasts with an 'extreme' form in which non-canonical subject coding reminiscent of SO-raising (e.g. accusative case marking) is, in fact, extended to matrix predicates that do not license accusative objects to begin with. Contrast the following examples from Wappo. In (78a), the (zero-) Accusative marking on the complement subject looks like a coding effect of SO-raising.[29] However, (78b) goes to show that this kind of non-canonical subject marking even appears where the NP in question cannot by any means be construed as a syntactic argument of the matrix predicate, e.g. in adverbial clauses. In Wappo, then, Accusative subjects have become a conventionalized rule for marking all kinds of dependent clauses:

(78) Wappo (Wappo-Yukian: USA; Thompson et al. 2006: 144, 154)

    a. *Ah*　　　*haṭis-khiʔ*　[*te-ø*　　　*takaʔ*　*mani-ya*].
       1SG.NOM　know-STAT　3SG-**ACC**　basket　carry-DUR.DEP
       'I know s/he is taking the basket.'

    b. [*I-ø*　　　*olol-o*　　　　*wen*]　*cephi*　　*waraha nayemi-seʔ*.
       1SG-**ACC**　dance-DUR.DEP　when　3SG.NOM　card　play-DUR
       'While I was dancing, she was playing cards.'

Similar cases of non-canonical subject marking beyond SO-raising are reported, by Arkadiev (2013), for Participial complements in Lithuanian (except for perception contexts, as we saw above), for Northern Uto-Aztecan and a number of Australian languages. Our sample language Kayardild is a well-known case in point, having a system of 'Complementizing case' that typically spreads to all constituents of the subordinate clause:

(79) Kayardild (Australian, Tangkic: Australia; Evans 1995: 512)

    Nyingka　*mungurru-wa*　[*ngumban-inja kajakaja-ntha*　*buka-nth*].
    2SG.NOM　know-NOM　　　your-**COBL**　　daddy-**COBL**　　dead-**COBL**
    'Do you know that your father's dead?'

Evans (1995: 542ff.) argues convincingly that the present-day system of subordinate case marking arose by reanalysis of an earlier system in which the subordinate

---

[29] If raising were involved, the position of the NP should also change, i.e. to the preverbal object position of *haṭis-khiʔ* 'know'.

constituents agreed in case with a main clause antecedent. This, then, relates to a stage in which the subordinate clause was adjoined to a 'head noun' or "anchor" (Dench and Evans 1988: 28) that was a proper argument of the matrix clause, not unlike in the perception-verb contexts from above. More generally, the historical evidence available across languages suggests that the three scenarios of non-canonical subject marking just discussed can form a diachronic cline (Fig. 4; the examples in line 2 are modelled after Lehmann's (2002: 59) historical discussion of Latin AcI constructions).

analogical extension > increasing lexical diffusion		
Petrus videt ACC*Paulum* INF*[currere].* 'Peter saw Paul [come].'	*Petrus vult* ACC*Paulum* INF*currere.* 'Peter wants Paul to come.'	INF*[Paulum currere] necesse est.* 'It is necessary [for Paul to come].'
matrix clause syntactically saturated and semantically entailed ('Peter saw Paul'), matrix object as "anchor"	matrix clause syntactically saturated but semantically not entailed (⊭ 'Peter wants Paul')	reanalysis of former matrix object as non-canonical subject of the subordinate clause = extension to matrix predicates that do not license the objects in question
control	raising	'quirky' subjects
e.g. Wambaya Infinitive*	Krongo, Motuna (cf. above)	e.g. Lithuanian (Arkadiev 2013: 418f.), Kayardild (cf. above), Wappo (no diachronic evidence)

Figure 4. Possible diachronic cline of non-canonical subject marking in subordinate clauses[30]

[*The Infinitive in Wambaya is intriguing in the present context; interested readers are invited to consult footnote 31.][31]

---

[30] It is, of course, also possible that the alleged stage 2 is skipped. On this pathway, the reanalysis of the former matrix object as the complement subject could also take place at stage 1, whereupon the pattern can be extended as a new construction. An intermediate stage 2 would be in keeping with small-scale extensions to only minimally different contexts first (i.e. those whose matrix verbs still license accusative objects) and then to more dissimilar ones (cf. de Smet's (2012) model of lexical diffusion). Likewise, it is possible that the development of raising constructions follows different pathways that do not involve the object-control pattern of stage 1. Givón (2001: 273) proposes that raising construction of type/stage 2 can evolve by blending a subject-control pattern (*I want to go*) with a simple sentence in which the same verb contains a lexical object (*I want him*), resulting in *I want him to go*. Therefore, Fig. 4 above represents but one possible mechanism for the rise of raising and quirky-subject constructions.

[31] The so-called "Infinitival" construction in Wambaya (Australian, West Barkly) is used in complementation "when the subordinate subject is coreferential with the main clause direct object" and "function[s] as adjunct modifying the […] object" (Nordlinger 1998: 213):

> *Ilinga-j-ba*   *nguyu-ny-u*   *gurla*   [*ngarl-warda*].
> hear-TH-FUT 3SG.NM.A-2P-FUT **2DU.ACC** talk-**INF**
> 'She will listen to the two of you talking.'

Crucially, the Infinitive has not penetrated any further into the complementation domain, and the matrix object has not been reanalysed as the subject of the Infinitive. Moreover, is intriguing to note that the Infinitival morpheme is synchronically also an agent nominalizer that is capable of forming headless relative clauses ('the one who is talking'). Therefore, it is at least a possibility that the structure above arose by placing the nominalization in apposition to the 'head noun' or 'anchor' in the matrix verb ('She will listen to you two, the ones who are talking'). This would be similar to the original stages in Lithuanian and Kayardild, where the current complement clause used to function as a modifier of an anchor in the main clause (cf. above).

Regardless of particular diachronic pathways, however, one can read Fig. 4 as a synchronic continuum of non-canonical marking of complement subjects. What the first two types have in common is that the internal subject is absent from the complement clause and expressed elsewhere. This situation contrasts with, and hence needs to be distinguished in the coding of the data from, so-called '**prolepsis**'. As Deutscher (2000: 57) points out, the two phenomena are similar but the term 'raising' "is normally used in the context of infinitival clauses ('I prove him to be a slave'), whereas 'prolepsis' is the traditional philological term used with finite clauses ('I prove him, that he is a slave')." As can be seen from these examples, the decisive structural difference between the two operations is that prolepsis leaves an overt 'copy' of the moved argument behind. For this reason, some specific instances of prolepsis have also been discussed under the rubric of 'copy raising' in the literature[32], or as a particular kind of "raising effect" again in Serdobolskaya's (2009) multivariate approach to raising. The following examples provide a flavour of typical proleptic structures:

(80) Slave (Na-Dene, Athapaskan: Canada; Rice 1989: 1245)

[ʔelá kʼínaretʼare ké kadįła   nį]   negháyeyidá   yíle.
 airplane   on **2SG**.go:out COMP   1SG.saw.**2SG**   NEG
'I did not see you get off the plane.'

(81) Jamul Tiipay (Hokan, Yuman: USA, Mexico; Miller 2001: 225)

a. [*Nyaach* *ø-iima-x*]=*pu*   *uuyaaw.*
    1SG.SBJ   **1SG**-dance-IRR=DEM   know
    'He knows that I will dance.'

b. [*Nyaach* *ø-iima-x*]=*pu*   *ny-u'yaaw.*
    1SG.SBJ   **1SG**-dance-IRR=DEM   **3>1**-know
    'He knows (about me) that I will dance.'

In (80) from Slave, the internal subject is indexed canonically on the complement verb but also appears as an object index on the CTP. This is similar in Jamul Tiipay, but here the contrast between a regular complement and one involving prolepsis also correlates with a difference in interpretation: The proleptic version "implies that the subject of the embedded clause has been a dancer for a long time" (Miller 2001: 225), i.e. it induces a habitual reading of the complement.

In contrast to raising, then, proleptic structures such as these do not result in non-canonical subject coding in my data: It would be odd to say that indexation on the complement verb is anything other than canonical here. However, there is a complication in some proleptic structures in the sample, which occurs when we find 'operator-satellite mismatches' again (cf. above), i.e. indexation going one way but NPs choosing another. For example, in Nkore-Kiga, finite complements can show raising of the internal subject NP to the direct-object position in the matrix clause; crucially, the complement subject is still indexed canonically, so there is some representation of it left inside the complement clause. This is illustrated in (82) below:

---

[32] Thus 'copy raising', as discussed mainly in the generative literature, refers primarily to sentences like *Jane seems like she is in a good mood*, i.e. in relation to specific matrix predicates such as *seem, appear, look, sound,* etc. Recent pertinent works include Potsdam and Runner 2001, Landau 2009, 2011, Asudeh and Toivonen 2012, and Kim 2014.

(82) Nkore-Kiga (Niger-Congo, Benue-Congo, Bantoid: Uganda; Taylor 1985: 80)

Ente z-aa-manya mukama waa-zo [*ku a-ri Wa-rubeba*].
cows 3PL-HOD-know **master of-them** COMP **3SG**-be Mr-Rat
'The cattle knew that their master was Rat.'

Given that "the conjunction *ku* is obligatorily moved to follow the [NP] *mukama waazo*" (ibid.), which in turn cannot be realized inside the complement, this is case of NP raising under retention of the canonical index. The structure is thus proleptic at a 'global' level, but since we have separate variables for the behaviour of NPs and indexes, we must take raising into account at a 'local' level. Very similar examples are found, for instance, in Skou, Gulf Arabic, To'aba'ita and Rama.

From a distributional point of view, it is worth pointing out that S-O raising effects and prolepsis are fairly common across the sample. Taken together, the two phenomena are found in (at least) 39 of the 100 sample languages. 9 of these involve prolepsis only, in such a way that the complement subject has a double representation by the same grammatical technique in both clauses (e.g. indexation, as in Slave above). For the other 30 languages, I had to code SO-raising effects in individual constructions because either a subject NP or a subject index was expressed in the matrix clause and not again inside the complement. However, there are two observations that delimit this number again: On the one hand, it must be noted that in 11 of the 30 languages in question, the raising effects were limited to perception verbs and, as we saw above, such contexts may better be analysed as involving control. Semantic contexts for which genuine raising could be postulated (e.g. on 'want' or 'believe') thus remain for 19 of the sample languages (as far as I can tell from the available evidence on major complementation patterns only). On the other hand, the raising effects found in those 19 languages are partly compensated for by leaving an alternative representation of the complement subject overt inside the subordinate clause. In other words, these cases involve 'local' effects of raising, but the global structure of the complement is proleptic (as in Nkore-Kiga from (82) above). This applies to 5 languages, and when these are subtracted, the number of languages with raising in the narrowest sense boils down to 14.[33] In view of this figure, we can fully agree with Noonan (2007: 82) in saying that "cross-linguistically, raising is not nearly as common as equi." For some of the sample languages, we have explicit statements that SO-raising constructions are unknown: In Warembori, for example, "[t]here is no possibility of the subordinate subject to be expressed on [the CTP] with object suffixes" (Donohue 1999a: 49). Similarly, "Lango does not allow subject-to-object raising" (Noonan 2007: 82). And for some languages, this statement extends to raising more generally, which is, for example, "unattested in Tzutujil" (Dayley 1985: 401) or Taba (Bowden 1997: 186). Finally, some radically isolating languages like Vietnamese and Chinese have structures which are indeterminate with regard to raising: Since the object clause is typically unmarked and simply inserted into the object position of the matrix, we end up with a structure in

---

[33] The languages with raising in the narrowest sense can be found in all macro areas except North America. There is a certain prominence of African and Eurasian languages, but more research is needed in order to substantiate this. In North America, it is not uncommon to find prolepsis instead of raising but again the overall figures are too small to warrant any conclusions.

which the complement subject often appears right after the matrix verb, i.e. in its direct-object position. Therefore, such constellations have been called "Kupplungssätze" (Nguyên 1979: 80 on Vietnamese) or "pivotal constructions" (Bisang 1992: 322), but, as Bisang himself admits (ibid.: 41), the evidence for assuming such a pivotal analysis is shaky since the structure is indistinguishable from an independent clause that, as a whole, comes to fill the object position of the matrix verb. For this reason, such cases were excluded in the above counts of raising languages.

From a theoretical point of view, raising and prolepsis have important implications for our analysis of complementation systems. In the literature, the two processes are often seen as signs of the "syntagmatic interweaving" (Lehmann 1988: 208) of the two clauses in a complex sentence, due to the explicit sharing of an argument. However, there is also one sense in which this leads to a syntactic *dissociation* of the clauses. In particular, the fact that an argument of the complement clause comes to occupy an argument position of the matrix predicate means that the complement clause itself is 'left behind', as it were: At least in those cases where raising and prolepsis lead to all argument positions of the matrix being saturated, the complement itself is deprived of the possibility of forming an argument of the CTP; consequently, it cannot be embedded as the subject or object in the main clause.[34] In the chapters to come, this effect has to be taken into account in several ways: First, it affects our statements on what kinds of syntactic function a given complement can fulfil: constructions that *always* involve SO-raising or prolepsis are precluded from functioning as direct objects in the language in question. Second, these constructions are of a different positional type than regular complements, i.e. adjoined to the matrix clause rather than being embedded in it. Finally, they can lead to a different classification of the pattern in question. Recall from §2.1 that one of Dixon's (2006a) definitional criteria for complement clauses is that they fill a syntactic argument position in the main clause. On a strict interpretation of this criterion, then, structures that *always* involve SO-raising or prolepsis processes on bivalent matrix predicates are complementation strategies, and no genuine complement clauses. As we shall see, there are not many patterns in the data that pose these wholesale problems. Normally, raising and prolepsis effects are restricted to certain predicate classes in which a given complement is used, so that the remaining contexts still provide unambiguous evidence for the genuine object status of the pattern in question. Nevertheless, they do affect the coding of the data at various levels, as I have just tried to show.

To round off the excursus on raising, let us finally have a brief look at how raising can affect the internal subject of subject clauses. From English, it is well-known that constructions like *It is likely* [*that Hannah will win*] and *It seems* [*that Julia is nervous today*] can raise their internal subjects to become the subject of the matrix clause. This produces the counterparts *Hannah is likely to win* and *Julia seems to be nervous today*, respectively. In a sense, SS-raising 'personalizes' the impersonal subject clause again. If

---

[34] In some syntactic theories, such as Relational Grammar (e.g. Perlmutter and Postal 1983), this insight is reflected by allocating the dependent clause in a raising construction a demoted syntactic function, the so-called *chômeur* function also applied to the demoted agent of passive constructions.

we focus on typical impersonal predicates such as *seem, likely, certain*, etc.[35], the cross-linguistic evidence for SS-raising is fairly scanty. In my sample, I have found explicit discussion of the issue in 7 languages, viz. Barasano, Basque, Karo Batak, German, Persian, Serbo-Croatian, Turkish, and another possible case in point in Vietnamese (cf. Nguyen 1979: 37f.). Since the phenomenon is only marginally relevant to the dissertation as a whole, I did not investigate the issue more systematically, e.g. with informants. Therefore, the figure given above is bound to be modified by future investigations, and I confine myself here to a few interesting qualitative observations.

For a start, let me provide an example of SS-raising (ex. (83b)) alternating with a regular impersonal subject clause ((83a)):

(83) Basque (isolate: Spain, France; Hualde and Ortiz de Urbina 2003: 653–654)

    a. *Ematen*   *du*   [*Jon*   *nekatuta*   *dago-ela*].
       give.IPFV   AUX   Jon.ABS   tire.PTCP   is-COMP
       'It seems that Jon is tired.'

    b. *Jon-ek*   *ematen*   *du*   [*nekatuta*   *dago-ela*].
       Jon-ERG   give.IPFV AUX   tire.PTCP   is-COMP
       'John seems to be tired.'

In example (83b), the internal subject of the complement raises to occupy the transitive-subject position of the matrix predicate, as indicated by the Ergative case suffix. In the following example from Karo Batak, SS-raising is an option for some impersonal predicates, and in the absence of case and indexation, it is achieved configurationally, i.e. by moving the complement subject in front of the matrix predicate:

(84) Karo Batak (Austronesian, Western-Malayo Polynesian, Sundic: Indonesia; Woollams 1996: 303)

    [*Kam*]   *banci*   [*berkat*].
     you   allowed   leave
    'It is allowed that you leave./You are allowed to leave.'

It is apparent from these examples that the syntactic status of the subordinate clause 'after raising' is still probably that of a complement and not an adjunct. In all cases of SS-raising that I am aware of, the matrix clauses do not appear to be impeccably grammatical without the rest of the complement. This contrasts with SO-raising, which despite resulting in a semantically different, i.e. non-entailed, proposition in the matrix clause, does not produce a matrix that is grammatically ill-formed if the complement is removed.

I close my discussion of SS-raising with a structure exemplifying a particular kind of 'raising effect' that Serdobolskaya (2009: 276) draws attention to: "the raised NP gets [a certain type of] marking, though the matrix verb does not have [the relevant] argument slot"; therefore, "a new syntactic position seems to be *created* especially for the raised

---

[35] In doing so, I neglect raising with phasal predicates, such as *The flowers began to wither*. In many languages where these constructions are possible, they arguably represent analogical extensions of phasal-P predicates (*I began to work*) to inanimate subjects. In other words, they are not typically the raised counterpart of an impersonal construction (e.g. *It began* [*that he worked*]).

NP." This happens in Barasano, where the verb *seyo* 'seem' is listed among the predicates that take a "sentential subject" (Jones and Jones 1991: 25). Crucially, this predicate never actually "has [a] subject of its own but pulls up the subject of the embedded clause (subject raising) and is followed by a pronoun referring to that subject." (ibid.: 161) This can be seen in the following example:

(85) Barasano (Tucanoan: Colombia; Jones and Jones 1991: 161)

[[*So*	*boha-ro-re*]	*bero-dẽ*	*bahi- ri*]	*seyo-ka-hʉ*	*ti.*
3SG.F	want-NMLZ-OBJ	like-SPAC	be-PTCP	seem-DST.PST-N3	**3INAN**

'It seemed that it happened (lit. 'was like') just as she wanted it to.'

In (85), the complement clause has an implicit inanimate subject ('it' (in 'it happened')), which comes to be expressed explicitly by a personal pronoun in the matrix clause. Note that Barasano is an OVS/VS language, and this is why the raised subject ends up sentence-finally. The motivation for this construction in Barasano is discourse-pragmatic, as it is used when "the speaker […] is emphasizing that the statement is from his viewpoint" (ibid.). This is what I referred to above as the 'personalization' of an impersonal predicate. From a syntactic point of view, however, the construction poses an interesting analytical challenge: Since *seyo* only occurs in the raising construction and does not license any other kind of subject, it is questionable whether Jones and Jones' analysis in terms of 'sentential subject' is justified for this predicate. Again, the subordinate clause is clearly a complement (it cannot be omitted), as is typical of SS-raising, but it cannot be a subject clause, not even an 'extraposed' one (since the matrix subject is not a placeholder, or 'dummy' subject, but a fully referential one). In fact, the construction works like an object clause in many respects, notably its referential matrix subject and its position in the sentence. For this reason, we will not assume that *seyo* involves a subject clause, despite SS-raising having occurred.

The purpose of our little excursion into raising patterns was to show that these have significant effects on the coding of the complement data. Most importantly, they were presented as the final 'special' type of treating the internal subject of complement clauses, and with this in mind, we can now return to the different-subject parameter as such. What we saw along the way is that there can be mismatches between indexation and case marking, or different treatments of the complement subject in subject and object clauses. To accommodate these, I decided to introduce a multivariate coding procedure. More specifically, the treatment of different subjects (henceforth DS) in complement clauses was broken down into the following variables and values:

- **Subject indexation in DS-contexts**. This captures the rules of DS-agreement for the complementation pattern as a whole; cases where subject and object clauses differ on this parameter are rare and will be taken care of by the other parameters below. The variable can take on the following values:

Neutral	The language does not have indexation as an inflectional category of the verb.
Canonical	The DS-index on the complement verb occurs in the same form as in

independent clauses.

Deletion	The DS-index cannot be overt on the complement verb.
Special	The rules for DS-indexation differ from that of independent clauses, e.g. a change to possessive indexes, replacement by relative indexation (switch-reference marking), use of a special subordinate indexation paradigm, or raising of the index to the matrix verb.
SS.only	The construction has not been found used in DS-contexts, neither in object clauses nor in subject clauses.
Combinations	Due mainly to the two types of coreference relations in DS-contexts (object control and disjunctive reference), various combinations of the above values can be attested for a complementation pattern that is used in both contexts. For exploratory purposes, the combinations are not reduced to a 'mixed' category but spelt out, e.g. 'canonical or special', 'special or deletion', etc.

From the theoretical perspective of de-/recategorization, a similar picture emerges as the one given for same-subject expression in the previous section:

Table 8. Numerical coding schema: Expression of DS-indexation[36]

1	Deletion, special coding, and mixed patterns with combinations of these values (= highly decategorizing and partially also recategorizing)
0.5	Mixed patterns with partially canonical expression types
0	Neutral and canonical expression (= no de-/re-categorization)

- **The form of different-subject NPs in P-clauses.** This captures the form of subject NPs in object clauses (including pronominal NPs where languages do not have indexation). The variable can take on the following values:

Canonical	The subject NP occurs in the same form as in independent clauses.
Free omission	The subject NP is normally omitted due to coreference with the matrix object, but this follows the rules of discourse anaphors in independent clauses and does not involve obligatory deletion processes (cf. same-subject processes above for the same category).
Deletion	The subject NP cannot be overt in the complement clause (and is usually controlled by a matrix object argument).
Special	The form of the subject NP is different from that of independent clauses, e.g. a change to a possessive construction, use of non-canonical case marking (recall Wappo from above), or raising of the subject NP into the matrix clause.
Indeterminate	The form of the subject NP (especially pronominal NPs) is the same for personal and the possessive NPs. Relevant only in certain nominalizations.
SS.only	The construction has not been found to be used as an object clause in DS-contexts, but only in SS-constellations.
Combinations	Same comments as above.

---

[36] Again, the 'SS.only' patterns will not be coded numerically, but will instead be left in their categorical form. In this way, they cannot increase the overall index of decategorization or nominality for a given complementation pattern.

From the perspective of de-/recategorization, the following picture emerges:

Table 9. Numerical coding schema: Expression of DS-NPs in P-clauses[37]

1	Deletion, special coding, and mixed patterns with combinations of these values (= highly decategorizing and partially also recategorizing)
0.5	Indeterminate, and mixed patterns with partially canonical expression types
0	Canonical expression or free omission (= no de-/re-categorization)

- **The form of different subjects in S-clauses**. This captures chiefly the form of subject NPs in S-clauses (including pronominal NPs where languages do not have indexation). However, where subject clauses show a special agreement behaviour not found in object clauses, this is incorporated here. For example, Hungarian Infinitives can show subject indexation only in subject clauses, but not when they are used as object clauses:

(86) Hungarian (Uralic, Finno-Ugric: Hungary)

    a. [*Haza*] *akar-ok* [*men-ni*].
       home want-1SG go-INF
       'I want to go home.' (Rebrus and Babarczy n.d.: 378)

    b. *Fontos*   *volt Péter-nek* [*olvas-ni-a*].
       important was Peter-DAT   read-INF-**3SG**
       'It was important for Peter to read.' (Kenesei et al. 1998: 35)[38]

The values distinguished for this variable are largely the same as in the previous category, except that 'deletion' is extended to 'generic-implicit' subjects, as discussed above. Therefore, the following numerical values apply:

Table 10. Numerical coding schema: Expression of DS-NPs in S- and A-clauses

1	Deleted and generic-implicit subjects, special coding, and mixed patterns with combinations of these values (= highly decategorizing and partially also recategorizing)
0.5	Indeterminate, and mixed patterns with partially canonical expression types
0	Canonical expression (= no de-/re-categorization)

- **The form of different subjects in A-clauses**. Exactly the same comments apply as to S-clauses, including the numerical coding scheme.

---

[37] Again, the 'SS.only' patterns will not be coded numerically, but will instead be left in their categorical form.

[38] Interestingly, "person-marked infinitives in [object] complement functions were acceptable in a previous period of the language" (Kenesei et al. 1998: 35). The present situation is likely to be a conventionalized usage-frequency effect, but I will not pursue this issue here (cf. Haspelmath 2008 for discussion of this type of frequency effect in diachrony).

At the end of this section, let me provide one example of what the above values look like 'in practice'. Nominalized complement clauses in Basque, which we have seen at several points in this chapter,

- generally remove their subject index but can show an SO-raising effect of the index in perception contexts (leading to 'deletion' on the same-subject-parameter and a mixed 'deletion or special' pattern on the different-subject-indexation parameter)
- leave the subject NPs in canonical form when they are used as object clauses, or show a raising effect again with perception verbs (leading to a mixed 'canonical or non-canonical' pattern on the DS-NP variable)
- leave the subject NPs in canonical form when they are used as subject clauses, but also exhibit implicit subjects of the coreferential and generic type.

This goes to show again just how complex the argument-structural configurations can be in a single complementation pattern, and they are, of course, not exhausted by looking at the internal subject of the complement clause. In addition, at least one further parameter needs to be taken into account, which we will turn to now.

4.4.2 The coding of the internal direct object[39]

The properties of the internal direct object were not coded in the same amount of detail as those of the subject: The parameter does not usually vary for subject and object clauses, and in the coding procedure, a simple two-way distinction between 'canonical' and 'non-canonical' coding patterns can be made. **Non-canonical object marking** is not a prominent pattern in the data, but it does occur in 28 complementation patterns ($N_{pat} = 28/228 = 12\%$), after all. These in turn distribute over 25 of the sample languages, from all macro areas.[40] In principle, the kinds of non-canonical object marking we find are similar to the ones distinguished for subjects above, though some additional patterns are also found:

- **Possessive coding**: Comrie (1976) and many typological publications ever since have observed that objects of action nominalizations are less prone to possessive coding than the corresponding subjects.[41] In other words, objects 'lag behind' subjects in the recategorization process. Where possessive marking does occur (resulting in the so-called POS-POS type of nominalization in Koptjevskaja-Tamm's (1993) study), it contributes to turning the complementation pattern into a strongly nominalized entity (i.e. a full-fledged NP rather than a clause). As was outlined in §2.3.2, possessive marking on the object is

---

[39] I concentrate on *direct* objects here since the form of other kinds of objects (e.g. oblique arguments) is typically the same in independent and dependent clauses, while the direct object is more likely to be affected by dependent coding (cf. Comrie and Thompson (2007: 355) on this point in the context of nominalization).

[40] Despite the fact that South American languages seem to be particularly prominent here, the proportion of languages with non-canonical object coding is quite similar in all macro areas and no significant areal skewing can be determined (randomized $\chi^2 = 4.11$, $p = 0.55$).

[41] Note that this asymmetry between subjects and objects is not confined to possessive marking, i.e. *re*-categorization processes, but is also (in fact, much more widely) found with regard to *de*-categorization. In Amele, for example, nominalized complement clauses lose their subject index but can retain canonical object indexation. Such asymmetrical coding patterns have led some researchers to posit a distinct cross-linguistic construction type of complement clause, viz. 'infinitives' (Noonan 2007: 67f.).

thus taken as a diagnostic for distinguishing complement clauses from complementation strategies. Representatives of such complementation strategies from the sample include nominalizations in Tamashek and Tukang Besi:

(87) Tamashek (Afro-Asiatic, Berber: Mali; Heath 2005: 676)

    Wằr   *ærhe-ɣ*   [*a-hænằy-ənnes*].
    NEG   want.PRFV.N-1SG.SBJ   SG-see.NMLZ-**3SG.POSS**
    'I don't want to see it.'

- **Special coding**: There are several types of special coding in the data: (i) **Case spreading** in Australian languages, such as illustrated for Kayardild in (79) above; as was stated there, such 'complementizing cases' often spread to all constituents of a complement clause and hence also affect the object. In Wambaya, similarly, the internal object of one complement appears in the Dative rather than the canonical Accusative case. (ii) The 'opposite' pattern is found when the direct object is deprived of the marker that it would normally take. This **loss of marking** is found in Semelai and Tariana; in the latter language, the object is also restricted to occur preverbally, which contrasts with a more flexible order in independent clauses. Such a **configurational change** of the direct object is also attested for one complement in Wari', where the object must be postposed behind the complement clause. (iii) Analogues of **logophoricity**: In Yagua, the internal object is marked by a special 'coreferential' marker if it is coreferential with the matrix subject ('They$_i$ said the storm would kill them-LOG$_i$'). (iv) In some languages, the direct object undergoes a change in status and form because the whole complement needs to be **detransitivized** in all or some contexts. This is found in West Greenlandic, Mosetén, Tzutujil and Yakan. An example follows from Mosetén, where the nominalizer *-dye'* can only be attached to intransitive predicates; if the complement is transitive, antipassivization needs to occur, so that the object appears in a demoted form:

(88) Mosetén (Mosetenan: Bolivia; Sakel 2004: 432)

    Yäe   *ködye-ye*   [*sob-a-k-dye'*   *öi-yä'*   *phen*].
    1SG   beg-1SG/2SG   visit-V-ANTIP-NMLZ   DEM.F-**ADESS**   woman
    'I beg you to visit this woman.' (lit. 'at this woman's')

A yet more extreme form of this is a complete ban on objects to occur; this is found in certain nominalizations in Mapudungun and Wari'.

For the sake of completeness, it should also be mentioned that our final type of special argument coding from above, i.e. **raising**, can also affect the internal object. The two varieties of raising in question are O-O-raising (Noonan 2007: 81) and O-S raising ('tough movement', cf. Berman 1974 for a classic study). O-O raising, where discussed explicitly, was taken into account as special object marking, but for O-S raising, Comrie and Matthews (1990) show that in-depth language-specific analyses are necessary to tease genuine object raising apart from superficially similar movement processes (e.g. topicalization in Chinese). But such analyses are clearly beyond the scope of the present paper. Therefore, O-S raising was neglected in the coding process.

Just as with the complement subject, the various processes of canonical and non-canonical marking can also combine in a single complementation pattern, resulting '**mixed**' coding properties of the internal object. Some reasons for such mixed treatments are listed in the following:

- **Referential properties of the object**: Sometimes, lexical objects are treated canonically, while pronominal ones can or must undergo certain changes (e.g. change of position in Noon Infinitives, possessive (*Ezafe*) coding in Persian nominalizations). In the Tzutujil Infinitive, definite or referentially specific objects require detransitivization and are hence changed in status and form.

- **Object properties may change depending on the presence and form of the complement subject**: The overt presence of a subject in Gulf Arabic nominalizations triggers oblique (PP-) coding of the direct object. In Tamashek nominalizations, the absence of the internal subject triggers possessive marking on the object, while the presence of the subject leads to the object being coded as a Dative PP. In To'aba'ita nominalizations, either the subject or the object is coded possessively, but not both; and if the subject takes possessive form, the object is either canonical or coded as a PP. Hence the overall pattern for objects is truly mixed.

- **Other reasons:** In Jamsay, the so-called "verbal noun clause" takes both subject and object arguments in possessive form, but this type of marking is suspended for object NPs if they are separated from the verb of the complement clause, e.g. by an intervening indirect object. In both Yakan and Tümpisa Shoshone, the objects of certain complementation patterns need to be equi-deleted under specific conditions.

This précis of the techniques of non-canonical object marking in complement clauses will suffice to convey the general impression, though the individual cases would doubtlessly be worth more detailed discussion. From the perspective of de-/re-categorization, the picture they produce is fairly clear:

Table 11. Numerical coding schema: Expression of the complement object

1	Non-canonical expression (= highly decategorizing and partially also recategorizing)
0.5	Mixed patterns of canonical and non-canonical expression types (= partial de-/re-categorization)
0	Canonical expression (= no de-/re-categorization)

4.4.3 Finding patterns of argument-structural coding

The detailed discussion and coding of argument-structural parameters has put us into a position to explore how the different types of subject and object marking in complement clauses combine in the data. This can be done on the basis of exploratory statistical methods. In order not to take us too far away from the overall thread of the present chapter, I will only provide a very brief, non-mathematical introduction to the methods to be employed, without any methodological discussion. I will then use them in a truly exploratory way to see if there are any trends in the argument-structural properties of our data and to detect noteworthy 'outliers'. Moreover, the analysis will have to be confined to object clauses only; since many complementation patterns in my data cannot be used in S- or A-function, they cannot be assigned a value on the subject-coding parameters for S- and A-clauses. The considerable amount of empty

cells that this leaves in our data is something that exploratory methods are not able to handle, and so subject clauses must be ignored in the present section.

We are left, then, with four categorical variables for the coding of the complement subject: the treatment of same-subjects, the form of indexation of different-subjects, the form of NPs of different-subjects, and the form of the direct object. In the following, I will make parallel use of two structure-detection methods which can inform us, in complementary ways, about the patterns that these four variables yield in the data. The first method, *Configural Frequency Analysis* (CFA, cf. von Eye 1990), basically outputs all possible permutations in which our variables and variable levels can occur and indicates their respective frequencies in the sample. Additionally, it implements an extended Chi-squared test in comparing the observed against the expected frequencies and assigns each combination in the data a corresponding χ^2-value, an (adjusted) *p*-value and an effect size measure (cf. Gries 2008: 242ff. for further information). The second method, the so-called *NeighborNet* technique (cf. Huson and Bryant 2006), is a particular kind of clustering algorithm that compares all of our complementation patterns to each other in order to examine which constructions behave similarly on the four argument-structural variables (cf. Cysouw 2007 and Bickel 2010 for the method and such and several applications to typological data). The result is a visual representation in the form of a 'split graph' or 'network'. The thrust of this visualization technique is nicely illustrated in Bickel (2010):
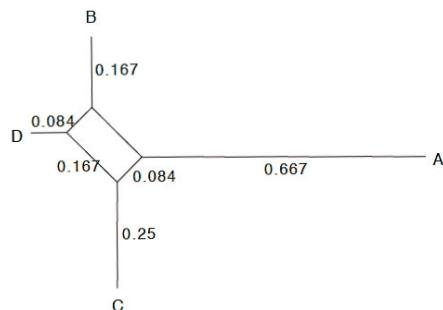
Figure 5. The logic of split graphs (taken from Bickel 2010: 83)

As one can see, the idea is to represent the relative similarity between the data points in geometrical space, and the total length of a connected line one has to pass to get from A to B in the graph reflects the underlying distance of the two data points: the less space to travel, so to speak, the more similar the data points. Dissimilarities in the data cause so-called 'splits' in the graph, more 'edges' to traverse and ultimately fewer clear-cut clusters of the data points.[42]

What do the two techniques reveal in relation to our argument-structural data? Turning to the CFA first[43], we learn that there are no fewer than **77** distinct

---

[42] The rationale and mathematical procedure behind the *NeighborNet* technique is laid out in Bryant and Moulton (2004). In this paper, the authors also explain its close relationship to better-known clustering methods and algorithms, such as 'neighbour joining' and the 'average linkage (UPGMA)' method.

[43] The analysis was performed by drawing on the script *HCFA 3.2* for *R*. I am grateful to Stefan Th. Gries for making this script available.

combinations (or patterns) of argument-structural organization that complement clauses exhibit. 44 of these are truly unique, being attested only once. This certainly shows how individually each complementation construction can be structured, and just how much cross-linguistic diversity there is. There are, however, also a few types that recur in the data, and some of them even with more than chance frequency. In particular, 59 of the 226 constructions (= 26.1%)[44] have the value 'canonical' on all four variables. These, then, are languages with indexation as the primary means for expressing pronominal subjects, where complement subjects retain the canonical index in all kinds of participant constellations, and where the internal object is not changed either. Complementation patterns of this kind, scoring the same on all relevant variables, are thus expected to cluster in a geometrical representation of the data. This can, indeed, be observed if we examine the *NeighborNet* graph fitted to our data.[45] Due to its size, I have put it in the Appendix (cf. Material 4). As can be seen, the graph structure is extremely complex, having to accommodate over 220 data points, but certain clusters can definitely be discerned. The strongest constellation identified by the CFA, i.e. the 59 canonical patterns from above, indeed all group together in the network; they appear at the bottom of the graph, in what looks like a densely populated (and hence dark) triangle pointing towards the ground. The relevant constructions are assembled in a long list extending downwards from this triangle; for ease of accessibility, I have highlighted them by grey shading. In the CFA, this cluster of constructions achieves the status of a so-called 'type', i.e. a statistically significant entity in the data ($p = 8.1e\text{-}16$). This configuration also has a very robust cross-linguistic distribution, being found in complementation patterns from all macro areas.

Another 'type' in the CFA sense is represented by languages without indexation, free subject-pronoun omission under coreferentiality, and fully canonical different-subjects and objects. This configuration occurs 15 times in the data and is significantly more frequent than expected by chance ($p = 0.002$). In the *NeighborNet*, the relevant constructions form a cluster on the middle left of the graph, again highlighted by grey shading. Corresponding languages include Tümpisa Shoshone, Malayalam, Vietnamese, Lao, Japanese, Imonda, Khwe, Epena Pedee and Hup, each of which possesses at least one representative of the complement type in question. Interestingly, most of the relevant structures are either paratactic clauses inserted into a matrix argument position, or are historically derived from quotative constructions.

Two other types that achieve significance are found at the other end of the coding spectrum: One has special coding for all types of subject but a canonical object (N = 4, $p = 0.005$); it applies to both complements in Evenki, and one construction each in Turkish and Quechua; all of them involve possessive coding of the subject. The cluster (or rather branch) they form in the network is highlighted again, on the middle right of

---

[44] Note that two of the original 228 data points had to be discarded for the present analyses since they are exclusively used as subject clauses but not as object clauses.

[45] The *NeighborNet* algorithm computes a distance matrix from our categorical data and plots the distances in geometrical space using split graphs. The algorithm is implemented in the software *SplitsTree4* <http://www.splitstree.org>. I am grateful to the programmers Daniel Huson and David Bryant for making this software and the manuals freely available.

the graph. The other significant type has deleted subjects in all environments in which it occurs and a non-canonical object (N = 3, p = 0.008); this is found in Semelai, Wambaya and Wari', which appear as a cluster at the top right corner of the *NeighborNet* (cf. the shading again). Interestingly, all three structures have independently been referred to as "infinitives" in the literature, despite the fact that the verbal morphology is rather different in each case.

In view of the many unique or only slightly different combinations in the data, it is not at all surprising that the CFA does not yield any other significant types (and that effect sizes are very low throughout). However, a few other combinations stand out by at least recurring in the data. They represent argument-structural patterns that are very similar to the 'types' from above but deviate on one of the variables. For example, variations on the largely 'canonical' patterns from above can be due to logophoricity in same-subject-contexts (N = 6), a constraint that the construction cannot be used in same-subject contexts (N = 9), or the occurrence of NP raising alongside canonical expression of different-subjects (N = 5). What we would expect, then, is that such similar patterns still form a cluster to some extent with the significant types they deviate from on a single variable only. This is borne out in the network display. For example, a bundle of constructions appearing in the lower right corner of the graph (e.g. from Skou, Nkore-Kiga, Arabic and Rama) clusters precisely those patterns that are identical in agreement properties with the biggest grey group, but which are proleptic in that the subject NP can be raised to matrix-object status while internal indexation remains the same.

The graph also shows a supra-division into more canonical and less canonical patterns, which basically represent the lower half and the upper half of the network, respectively. Not surprisingly, therefore, the majority of the argument-structural configurations in the upper half tend to co-occur with dependent verb forms, while the ones in the lower half typically occur in 'finite' constructions with independent verb forms (Fisher exact test, p = 2.2e-16, odds ratio = 22.33). As can be seen, the upper half of the network has a very diverse internal structure, with many individual branches rather than a few clear-cut 'edges'. There is, however, a principled pattern motivating the extreme 'outliers' in this area (and everywhere else, incidentally): The extremely long branches reaching out to the margins of the graph deviate from the clusters they belong to by showing non-canonical coding of the complement object. The network thus reflects, for example, the pattern of 'case spreading' in Kayardild ("CompKay" in the upper left of the graph), as well as the fact that the two extreme outliers on the upper right ("CompMap1" and "CompWar2") are precisely the two constructions in the data that do not allow an internal object to begin with (cf. §4.4.2 above). As a final example, let us look at the lower right corner, where the long branch sticking out for Yagua ("CompYag1") marks a deviation from the other languages in the same cluster: Unlike Krongo, Lango, West Greenlandic, Fongbe, Jamsay and Kana, the construction in Yagua shows logophoric coding not just for subjects, but also for objects.

This will be sufficient to convey the general idea and the usefulness of these clustering techniques for detecting groups as well as individuals in the data. We will

return to such methods later on in the dissertation, and we are also leaving behind now the topic of argument structure more generally.

4.5 Boundary-marking devices and nominal flagging

The last parameter to be introduced in this chapter relates to the outermost layer of the complement clause, i.e. to specific grammatical marking at the clause boundary. This topic has been investigated in linguistic typology from a variety of different perspectives. Several lines of research have emphasized the importance of overt boundary markers for parsing the division between matrix and subordinate clauses (cf. Hawkins 1994 for a specific theory, and Dryer 2009 and Heath 2010 for the more general relevance of boundary marking in complex sentences). Other researchers have foregrounded the semantic impact of boundary-marking devices in complementation. From this angle, complementizers and related morphemes may contribute significantly to the modal orientation of the complement as a whole (cf. Ransom 1986 and Frajzyngier 1995, to name but two representatives), and they also serve, of course, to establish a particular semantic relationship to the matrix clause: Specifically, they can signal, or 'flag', that the complement is to be interpreted as a certain kind of argument of the matrix verb. In this function, boundary-marking devices underscore the nominalization of the complement clause: The more argument-like or NP-like the complement is made, the more advanced the process of de-/re-categorization. It is chiefly in this latter context that boundary-marking devices are relevant to the present chapter; they constitute the final parameter to contribute to the overall degree of desententialization of complement clauses.[46]

The 'practical' relevance of this parameter becomes apparent when we recall the cases of so-called 'clausal nominalization' mentioned in §4.2 above. Specifically, in (35) we encountered an example from Jamul Tiipay, which is repeated here for convenience:

(89) Jamul Tiipay (Hokan, Yuman: USA, Mexico; Miller 2001: 224)

Puu-ch   [*nya'wach neyiw-x*]=*pu*   *map.*
that.one-SBJ   we.SBJ   come.PL-IRR=DEM   want
'He wants us to come over.'

In this example, a full-fledged clause without any clause-internal sign of desententialization can function as a complement clause if it is integrated with the matrix clause by means of a demonstrative clitic. This clitic helps to signal the boundary of the complement clause, which is particularly beneficial here given that the complement is centre-embedded between two constituents of the matrix clause. The demonstrative does, however, also arguably signal the nominal character of the

---

[46] In the process of coding the data, I recorded information on the form, the position, the omissibility and – where retrievable – on the etymology of the boundary markers that characterize each complementation pattern in the sample. This information is collected in the database, which the reader is kindly referred to for a general impression. Etymological aspects will concern us chiefly in Chapter 7 of the dissertation, while the interaction of positional, formal and diachronic properties of subordinators is studied in the larger project with which the present work is affiliated. Therefore, the present section will concentrate exclusively on their nominalizing dimension.

subordinate clause in question; accordingly, complement clauses in Jamul Tiipay are referred to as "nominalizations" throughout Miller's (2001) description of the language, despite the absence of 'lexical nominalization' of the predicate. Clearly, then, there is a sense in which complementation patterns like (89) above are 'more nominal' than a similarly finite clause that does not have a subordinator at all, or one that is not associated with nominal morphology, such as a quotative marker or an adverbial conjunction. In the present section, therefore, I will investigate to what extent the complementation patterns in the sample exhibit nominal flagging at the clause boundaries, i.e. either as the only means of nominalizing the complement or in addition to the lexical nominalizers or other dependent morphology on the verb.

All complementation patterns in the data were scanned for such nominal flags. In roughly two thirds of all cases (N_{pat} = 154/228 = 68%), no such marker is present (or no marker in addition to a dependent verb form). The remaining cases have nominal morphology of various sorts. Table 12 below provides an overview of these types and indicates how frequently each flagging type occurs in addition to dependent morphology on the verb:

Table 12. Nominal flagging by boundary markers of the complement

Type of nominal marking	F(dependent verbs)	F(independent verbs)	F(abs)
Case marker	20	4	**24**
Determiner	10	9	**19**
Adposition	5	2	**7**
Clausal nominalizer	4	1	**5**
Gender or noun-class marker	2	1	**3**
Grammaticalized head noun	1	2	**3**
Topic marker	2	0	**2**
Several at once	6	5	**11**
Total	50	24	**74**

As can be seen, the most frequent types of nominal flagging come from case markers and determiners. The former are typically added to lexical nominalizations (or related verb forms), while the latter co-occur with dependent and independent forms with about equal frequency (the difference between case markers and determiners is mildly significant in a Fisher exact test, p = 0.04). The combination of **determiners** with independent verb forms typically yields structures such as the one from Jamul Tiipay above, which are also found, for example, in Amele, Chumash, Jamsay, Korafe, Motuna, Tzutujil and Yagua. **Case markers** as a category are fairly straightforward. Note, however, that the above counts for case markers only include non-zero morphemes, i.e. overtly signalled cases. If a complement clause occurs in S- and P-function only, and these receive a zero-Absolutive marker on NPs, no case marking was recorded for the complement. An illustration of the kind of case marking noted in Table 12 is given from Imonda below, where a nominalized clause can receive the polyfunctional 'goal/object' case marker -*m* in complementation:

(90)   Imonda (Border: PNG; Seiler 1985: 84)

[*Ièf   sabla nibia-ual-l-m*]   *õ-f.*
 house two   build-DU-NMLZ-**GL**   say-PRS

'I want to build two houses.'

Sometimes the occurrence or non-occurrence of a case marker with a particular kind of complement is noteworthy from the perspective of the case-marking system as a whole. In Malayalam, for example, Accusative coding generally follows a pattern of so-called 'differential object marking' (Comrie 1989: 129ff.) such that only objects high in animacy normally receive the Accusative case, while inanimate entities do not. The fact that certain nominalized complementation patterns can, in principle, also take this marker, appears to be anomalous because they are, by definition, inanimate. It is thus possible that their occurrence is motivated by a desire to signal the integration of the complement into the main clause, i.e. in terms of boundary-marking and transparency of the linkage relation. For Wappo, Thompson et al. (2006: 140) point out as noteworthy that the so-called Infinitive (which is probably a lexical nominalization by origin) cannot get the appropriate case marker when it functions as a subject clause, while headless-relative clauses as subjects can. This specific pattern does, however, have analogues in other languages that could help to explain it. First, so-called 'Infinitives' have often been derived historically from case-marked nominalizations (cf. §4.2 above), and it seems to be common for these structures to fossilize in this form and not to take case marking again. Thus in German and Serbo-Croatian, for example, Infinitival complements are not Accusative-marked in object function, even though the corresponding object NPs with the same verbs require this type of marking. Second, there is a cross-linguistic pattern for nominalizations to be case-marked when they function as headless-relative clauses (or 'product' nominalizations) while the same nominalizing morpheme is not accompanied by a case marker if it has a complement (or 'process') interpretation. In our sample, this is reported, for instance, for certain propositional-attitude verbs in Burmese (Soe 1999: 315). While headless relative clauses in Wappo involve morphology that is distinct from the Infinitive, the differential marking pattern with regard to case is similar. Therefore, the peculiarities of case flagging in Wappo Infinitival complements may ultimately be explicable by taking these two cross-linguistic observations into account.

Adpositional flagging is similar to case marking but often semantically more specific (e.g. German *zu* 'to', Amele *nu* 'for', Tzutujil *chi* 'at, to, with'); adpositions such as these correspond to bound oblique case markers in other languages (e.g. Instrumental *=s* in Tamashek). The category of '**clausal nominalizers**' refers to morphemes that have scope over an entire clause rather than the verb, and are not determiners but genuine nominalizers in their own right. They include, for instance, Amele *=ec/oc*, the "complementizer" *ní* in Slave and the versatile subordinator *xa=* in Chalcatongo Mixtex (cf. Hollenbach 1995 for its analysis as a clausal nominalizer). **Gender- and noun-class marking** is found, for example, in Yimas, where nominalized complements "are like nouns and nominals generally in being affixed with suffixes to mark noun class" (Foley 1991: 384); interestingly, these are specifically adapted to the respective CTP context, such as a 'Customary action' suffix for generic subject clauses, or a 'Desiderative' class marker, etc. In Yuchi, by contrast, complement clauses are not lexically nominalized, but the finite complements "have markers that flag them as

dependent clauses." (Linn 2001: 494) The most common one is the noun class clitic =*ci*, illustrated in (91) below:

(91)　Yuchi (isolate: USA; Linn 2001: 320)

[*Nedze-di-k'ã-'wede*]=*ci*　　　　*sẽ-:le*　　*lœ.*
　2SG.P-1SG.AGT-COM-talk=**SUB.CLF:SIT**　good-VERY　ENC
'It's been really good to talk to you.'

Noun-class markers typically derive from lexical nouns historically (this is very transparent in Yimas, cf. Foley 1991: 385), and this also holds for another type of nominal flagging, called '**grammaticalized head noun**' here. Thus certain 'complementizers' in Dolakha Newar, Modern Khwe and Ainu are evidently nouns ('thing/place', 'talk/matter/news' and 'manner', respectively). It is thus no accident that the complements in question are also commonly analysed as 'clausal nominalizations', so they will receive this treatment in the present paper as well. **Topic marking**, finally, has been found as a typical boundary marker of certain complements in Awa Pit and Korafe, as in the following example:

(92)　Korafe (Trans-New Guinea, Binanderean: PNG; Farr 1999: 79)

[*Mary*　*amb-ari*]=*mo*　　　*John*　*jo*　　*tumond-ae=ri.*
Mary　die(I)-NMLZ=**TOP**　John　NEG　believe-not.do=COP.ASS
'John did not believe that Mary died.'

As can be seen in Table 12, a combination of **several** nominal flags is also attested. In Wolaytta, for example, one kind of complement clause is marked by a dependent though inflected verb form, to which both a nominalizer and, in appropriate circumstances, a case marker are added (Wakasa 2008: 229):

(93)　Wolaytta (Afro-Asiatic, Omotic: Ethiopia; Wakasa 2008: 186)

Táání　[*ʔí*　*kais-ó*　　*gid-ídoo-g-áá*]　　　　　　　*yoot-áas.*
I　　　he　thief-ABS　become-REL.PRFV.NSBJ-**NMLZ-ABS**　tell-PRFV.1SG
'I told that he was a thief.'

Nominalized complement clauses in Basque are formed by "adding to the verb-stem the gerund suffix -*te* or -*tze* followed by the definite article -*a*. The resulting NP can appear in any NP position in a sentence and must take any case-marking appropriate to its grammatical position." (Hualde and Ortiz de Urbina 2003: 167); the nominalizing morpheme is thus augmented by both a determiner and a case marker. Japanese *koto*-complements transparently contain the grammaticalized head noun *koto* 'thing, fact' as a subordinator and are additionally marked for the relevant case. In Menya, so-called 'dependent final clauses' are relatively more autarkic in terms of their inflectional behaviour than other complements in the language, but they carry a number of 'external' dependent markers. In particular, "they generally bear the same definiteness, personalizing clitics and case markings that equivalent nominals do" (Whitehead 2004: 192). And so on.

As with the other parameters in this chapter, we will include in the survey of nominal flagging some distributional and theoretical aspects of the phenomenon. From a distributional point of view, nominal boundary marking does not have particular areal biases; in almost all macro areas, it appears in 30%–45% of the relevant

complementation patterns. The only outlier here is the area of Southeast Asia and Oceania (SEAO), where such marking is much less common ($N_{pat} = 4/32 = 12.5\%$). A statistical comparison of all macro areas thus only achieves a significant skewing if SEAO is included in the calculations; if it is taken out, no difference between the remaining macro areas can be detected (randomized $\chi^2 = 3.9$, $p = 0.42$). In geographical terms, then, devices of nominal flagging largely follow the distribution of lexical nominalization in complementation (cf. §4.2 again).

A distributional characteristic of a very different kind which I would like to note relates to the obligatoriness (versus omissibility) of the flagging material in individual languages. Most of the nominal flagging we have looked at in this section appears to be an obligatory part of the complementation pattern that it marks. Variability of occurrence is noted explicitly for 20 of the 74 markers (= 27%), and the factors governing the choice are not often laid out. However, in coding the data, I came across some recurrent patterns of omissibility that would be worth investigating in a more principled fashion.[47] Let me mention just a few of them here.

One scenario is for boundary marking to be absent under subject coreferentiality in the two clauses and to be present under different-subject conditions. The clausal nominalizer *xa=* in Chalcatongo Mixtec follows this pattern rather clearly, although not without exceptions (cf. Macaulay 1996: 154). This split according to differential argument sharing between the two clauses is, of course, not confined to nominal boundary markers, but a frequently encountered effect for subordinators of all sorts. Thus in Gulf Arabic, the complementizer *ʔin(n)* is not frequent in colloquial speech to begin with and hardly ever occurs with certain "verbs expressing emotions and desires", but crucially, "where the subject of the embedded clause is not the same as the main clause, [the] complementizer must be used, even with [emotive and desiderative] verbs" (Holes 1990: 20). Similarly, finite complement clauses in Mosetén only show an explicit conjunction in different-subject contexts (if at all), but leave same-subject contexts unmarked (Sakel 2004: 429). Such usage preferences can also be(come) conventionalized coding patterns; in Choctaw, a certain type of complement has a strict same-subject constraint and importantly, it is the only type of complement clause in the language that does not have a subordinator (Broadwell 2006: 281). In the literature, the coding differences between same-subject and different-subject constellations have often been treated as an iconic reflection of the relative conceptual integration of the two events (e.g. Givón 1980, Cristofaro 2003). For desiderative complements, Haspelmath (to appear) proposes an alternative explanation in terms of differential usage frequencies and concomitant economical coding of the more frequent same-subject scenario (cf. also Zipf 1935, Croft 2003: Ch.4 for the relationship between frequency and economical coding). Although I am highly sympathetic to this

---

[47] A very recent attempt at such a more principled framework for understanding optionality in grammatical coding has been presented by McGregor (2013). His approach places problem-solving in communication centre-stage, with notions such as joint attention and interpersonal layers of grammar figuring prominently. No attempt will be made here to investigate to what extent the relevant data from the present sample provide evidence for this new theory. But more generally, I concur with McGregor in stating that optionality is a severely understudied aspect of grammatical structure from a cross-linguistic point of view.

latter line of argumentation, it remains to be explored by future research which role such lexically-specific (e.g. 'want') coding asymmetries play in the formation of more general or even system-wide differences between same-subject and different-subject constellations.

There are more cases of nominal flags in the data whose omissibility obeys a lexically specific constraint. For example, the demonstrative determiner flagging complements in Barbareño Chumash is obligatory except in indirect speech, where its occurrence is highly variable (Wash 2001: 89).[48] Similarly, the definite article in Lakota is apparently always present as a boundary marker with the CTP 'know', but more readily omissible with perception verbs and 'believe' (Buechel 1939: 227ff., Van Valin 1977: 71ff.). Finally, there is a very interesting lexically-specific effect of nominal flagging in the so-called 'Infinitive' in Turkish. The Infinitival morpheme *-mAK* is the control "variant" (Kornfilt 1997: 51) of the more general nominalizer *-mA*; they have co-existed since at least Old Turkic (Erdan 2004: 279, Hennesy and Givón 2001). Crucially, while complements in *-mA* are regularly flagged by Accusative case when they appear in object function, there is more variability with the Infinitival variant: "When an infinitive is the object of a transitive verb, the objective suffix is normally optional. As a general rule, the suffix is omitted if the main verb is one that is frequently used in this construction […], but retained if the main verb is less common." (Underhill 1976: 310) Among those high-frequent items, according to Underhill, we find 'want', 'like' and 'know (how to)'. What we observe here, then, is a lexically-specific frequency effect: The more a complement clause is predictable with a given CTP (and vice versa), the less often explicit boundary marking tends to be given. Note that we are really dealing with the marking of the boundary between the two clauses here since the Infinitive in Turkish precedes its associated CTP:

(94) Turkish (Altaic, Turkic: Turkey; Göksel and Kerslake 2005: 413)

    [*Sokağa+çık-mak*]-ø   *isti-yor-um.*
     go.out-INF-ACC   want-IMFV-1SG
     'I want to go out.'

It thus appears that by the time the parser encounters the Infinitival morpheme, the subsequent occurrence of *isti* is sufficiently predictable in stochastic terms. Given that the "surprisal" or "information value" (Levy and Jaeger 2007) of the following unit is low, no extra marker has to be inserted to make the relationship between the Infinitive and the CTP more explicit.[49]

One other case of omissibility is worth mentioning because it demonstrates an efficient interplay between overt flagging of the complement and its position in the

---

[48] The omissibility of subordinators in indirect speech is an extremely widespread pattern in my data. However, in most cases that I am aware of, this is because the subordinator of the utterance complement was itself derived from a quotative verb, such as 'say', which is synchronically still in use. Consequently, the subordinator is not employed if the CTP is the same verb again (cf. Rohdenburg's (2003) 'horror aequi' principle, or Menn and MacWhinney's (1984) 'repeated morph constraint'), even if the exact morphological shape of the two items may be different. In Chumash, by contrast, the situation is rather different since the subordinator is not a quotative marker.

[49] For this information-theoretic account to syntactic reduction processes, cf. Levy and Jaeger 2007 and especially Jaeger 2010 in relation to complementizer omission in English.

sentence. Yuchi used to have a fairly rich system of explicit marking of subordinate clauses, including the noun-class marker we encountered in (91) above as well as dependent tense and negative morphemes (Linn 2001: 495). This elaborate system of subordinate marking has fallen out of favour, to the effect that all of the dependent morphology can be dispensed with altogether (including the boundary marker). Crucially, however, this economical coding behaviour requires that the complement appear in the preverbal direct-object position of the main clause; if overt marking is present, speakers have the possibility of extraposing the complement to the right:

(95) Yuchi (isolate: USA; Linn 2001: 495, 497)

    a. multiply marked complement in postverbal position

      Wahe-'ē   *hi-'yo-'õda*   [*hǣ-syo-thla*   *shẽ=ci*]?

      why-ACT   3SG:INAN.P-2SG.AGT+'YU-know   **NEG.SUB**-3SG(EFM).AGT.VAL-go **PST.SUB=CLF:SIT**

      'How do you know that she didn't go?'

    b. unmarked complement in preverbal position

      [*Hẽ-tishǣ*   *jẽ-fa*]   *hi-dzǣ-le.*

      3SG(EM).AGT-lie   PST-PRFV   3SG:INAN.P-think-ITER

      'I found out that he had lied to me.'

It appears, therefore, that economical coding comes at the expense of positional freedom: If the dependent character and the function of the subordinate clause are not overtly signalled by the morphology, then the embedding in object position needs to compensate for that. Conversely, one could also say that the positional cue is a sufficient signal for assigning the function of the embedded clause; therefore, it allows for the morphological marking to be dispensed with. In this light, the pattern in (95) above would be another example of the relationship between predictability and efficient coding. As Hawkins (2004: 48) puts it, "the processor avoids the articulation and processing of explicit linguistic material that is derivable or inferable from the linguistic or larger context."[50] Therefore, it appears to me that predictability-based accounts of omissibility present a fruitful framework (perhaps in conjunction with McGregor's communicative one) for further exploring the phenomena of differential complement flagging noted here.

At the very end of the section, we can now draw up the numerical coding scheme for nominal flagging devices. All of the markers discussed here are recategorizing devices by definition. However, rather than using a binary "0/1" distinction, we can recognize a certain internal differentiation again: The highest degree of recategorization is displayed by constructions that have several nominal markers at once. As we shall see shortly in §4.6, there are languages in which two complementation patterns both involve nominal flagging, but they differ in the amount of nominal categories that are expressed (e.g. case only versus case, definiteness, noun class, etc.). For this reason, the highest value "1" in our coding scheme is reserved for such strongly nominalizing patterns. The ones showing one nominal element should not be treated *too* differently, though. Thus, I decided to code

---

[50] A superficially similar phenomenon is also found in Korafe, where complement clauses "can occur without a complementizer when they are phonologically set apart from the rest of the sentence" (Farr 1999: 78). However, it is not quite clear if this could be explained along similar lines.

the occurrence of the remaining categories as having a value of "0.85". This figure is arbitrary and could be debated, but given the number of other parameters that enter the calculation of the overall numerical index for each construction, it does not matter greatly whether we choose 0.85, 0.9 or 0.75. The coding schema for nominal flagging is set out in Table 13:

Table 13. Numerical coding schema: Degree of nominal flagging in complementation patterns

1	Several nominal flags at once (= multiple recategorization)
0.85	One nominal flag (= singular recategorization)
0	No nominal flagging (= no recategorization)

With all the parameters introduced and coded in such a numerical way, we can now proceed to a global look at the data.

4.6 Degrees of de-/re-categorization in complementation systems

In the preceding sections, I have provided a qualitative and quantitative overview of the internal structure of complementation patterns. The multivariate breakdown of the relevant structures makes it possible to 'locate' each complementation pattern in the data precisely on a scale of desentialization, understood in Lehmann's (1988) sense as reflecting processes of both decategorization and recategorization. Such a numerical scale was developed for all individual parameters, and their addition now leads to a **cumulative index of desentialization (CID)** for each data point. As an example of this calculation, let us pick out a complementation pattern from Yuracaré, exemplified in (96) below:

(96) Yuracaré (isolate: Bolivia; van Gijn 2006: 294)

Mi-bëjti   [*ayaj   tütü   mala-m=ti*].
2SG-see:1SG.SBJ   fast   sit;be go.SG-2SG.SBJ=SUB
'I saw you running.'

This complementation pattern, which can be used with a variety of different CTP classes, has the following internal properties (with 'ID' symbolizing 'index of desentialization' for every variable):

- The verb in the complement clause is an 'independent' form (ID = 0).
- The predicate bears regular TAM inflection, but some CTP contexts require the occurrence of the 'Intentional' morpheme, "which functions as a subjunctive marker here" (ibid.: 319). This amounts to what we called 'differentially special' TAM marking (ID = 0.25).
- The argument structure in object clauses is non-reduced in all respects. Indexation is the primary means for expressing subjects, and the index for the complement subject is overt in all constellations ($ID_{same-subject}$ = 0, $ID_{DS-indexation}$ = 0). (Note that, in (96) above, prolepsis has taken place (the complement subject being represented again as the object index on the CTP), but since the subject index inside the complement is never absent, this pattern does not result in any non-canonical marking.) The form of different-

subject NPs in object clauses is canonical (ID = 0).[51] The internal direct object argument is not changed in its form either (ID = 0).

- Manner satellites are not changed into adjectival form but remain adverbial (ID = o).

- The complement is obligatorily marked by the subordinating particle =*ti*, which normally occurs at the clause boundary. Crucially, this marker is a nominal flag: "The morpheme =*ti* [is] a nominalizer, not of the predicate itself, but of the entire clause. A possible analysis is that =*ti* is a pronominal deictic element, also found in the demonstrative *ati*." (ibid.: 313) Since this is the only trace of nominal flagging, the relevant ID = 0.85.

- This results in a **cumulative index of desententialization (CID)** of 0.25 + 0.85 = 1.10, making it a relatively weakly desententialized pattern.

The CIDs obtained in this way will be crucial in later chapters for identifying correlations between the internal structure of complement clauses and other aspects of their grammatical behaviour. In the present section, the CID values can be used for a different purpose, i.e. to gain a more global perspective on desententialization processes in the data. Overall, the complementation patterns in the sample show an average degree of desententialization of $CID_{mean} = 2.22$ ($SD = 2.18$). The range that the individual CID values span is rather wide: Some patterns in the data do not show a single trace of de-/re-categorization on the parameters in question (CID = 0), while the most extremely nominalized constructions go up to a CID of 7.85 (e.g. the deverbal noun constructions in Mapudungun and Georgian). On a higher level, the individual CID calculations also enable us to look at the organization, or internal structure, of complementation *systems*. To this end, I have calculated (i) the mean CID for each language in the sample, i.e. the average degree of desententialization that is reflected by its complementation patterns; (ii) the CID range for each language (i.e. the difference between the highest and lowest CID value), reflecting how structurally dissimilar the complementation patterns are within a single language. In Kolyma Yukaghir, for example, both complementation patterns in the data are relatively strongly desententialized, resulting in an average CID of 5.68 but a range of only 0.35 since the patterns are structurally similar. In Basque, by contrast, a sentential and a nominalized complementation pattern polarize more strongly, resulting in an average CID of (0.25+5.5)/2 = 2.88 and a fairly wide range of 5.25. From a typological point of view, it would now be interesting to see which complementation systems are organized in similar ways with regard to the mean and range of their constructions' CIDs. This question can be answered by turning to appropriate dissimilarity measures again. Specifically, the CID mean and CID range for each language can be used as variables to

---

[51] Importantly, the ID calculation for the form of subject NPs can only take object clauses into account, while the same parameter for S- and A-clauses must be disregarded in the CID. This is necessary, as a general policy, because of an inherent asymmetry in the data: Almost all complementation patterns in the data can be used as object clauses, but only a restricted subset is available as S- or A-clause (cf. Chapter 6 for detailed discussion of this point). Therefore, the constructions that *are* attested in subject function would automatically increase their CID if the argument-structural parameters specific to S- and A-clauses were taken into account, while all other constructions would receive lower CIDs. The resulting indexes would distort the picture asymmetrically, and hence the S- and A-parameters were not considered for the CID calculation. We will only take them on board again when analyses are performed on S- and A-clauses individually, i.e. without comparing them to P-clauses (e.g. in Chapter 5).

create a distance matrix of the data, which in turn can be submitted to an appropriate clustering algorithm.[52] I used the *NeighborNet* technique again to plot the results as a split graph. Just as in §4.4 above, this network-like display reflects the actual distances between the data points in geometrical space and is thus often more easily accessible than the output of a hierarchical cluster analysis with 100 data points.[53] The network is again found in the Appendix (Material 5).

As we can see, there are at least three extreme poles in the data. At the left end of the graph, we find languages like Mapudungun, Georgian, To'aba'ita and Tukang Besi. These languages have in common that their complementation systems polarize between a highly sentential and a highly nominalized construction. Thus they show similar CIDs on average and, more importantly, a similarly wide CID range. At the other ends of the network, both on the far right and towards the bottom, we find more homogenous complementation systems, which split into two classes again. On the far right, a whole list of languages is given which have identically low values on both the CID mean and the CID range. In these languages, then, the primary means for complementation are fully sentential, while a strongly reduced type of complement is missing. As was discussed in §4.1 above, this may be for two reasons: One is a potential sampling problem, in the sense that a nominalized construction could not be taken into account due to a lack of information. However, in the cluster of languages on the far right, complementation was described in the sources entirely without reference to nominalization processes, and so it is likely that these processes do, indeed, not play a comparably important role as in other languages, even if they do exist. In Abun (West Papuan, North-Central Bird's Head), for example, a lexical nominalization is not unknown, but it is only attested as a 'product' nominalization and not with the 'process' reading that would be necessary for complementation. Moreover, in Berry and Berry's (1999) discussion of complementation, which is very detailed and sophisticated, no mention is made of lexical nominalization. One is led to conclude, therefore, that this process is of limited or even no relevance to complementation as defined in the present study. A very similar situation is found in Mekens (Tupian, Tupari: Brazil). In this language, a productive lexical nominalization is a 'participant' nominalization for deriving instruments or locations (cf. Galucio 2001: 101ff.). In the domain of complex sentences, this can be exploited for causal clauses, but not for complementation, as far as I can see. In §4.1 above, I presented further languages for which the limited nature of lexical nominalization is explicitly argued for. In addition to those, let me quote from the description of Taba (Austronesian, Eastern-Malayo Polynesian, South Halmahera-West New Guinea: Indonesia):

> In many languages, the structure of complement clauses can be quite different
> from the structure of main clauses, e.g. with infinitival complement clauses or

---

[52] Since both variables are interval-scaled, I used Euclidian distance as a dissimilarity measure, implemented in the function `dist` in *R*.

[53] On the more general drawbacks of enforcing hierarchical clustering algorithms on typological data, cf. Cysouw 2007. For comparative purposes, I also performed such a cluster analysis (with 'average' linking) on the data; it yields essentially the same results as the *NeighborNet* algorithm, showing the close relationship between the two methods.

complement clauses whose subjects are 'raised' to the object position of the main clause. However, in Taba all complement clauses occur in the same form that they would be expected to take as main clauses. There are no non-finite clauses in these constructions." (Bowden 2001: 387)

In many of the languages on the right end of the network, then, complementation is rendered by fully clausal structures. Some of them simply insert an unmarked clause into an argument position of the main clause (e.g. Chinese, Vietnamese, Epena Pedee, Warembori) or adjoin it to a full-fledged main clause (e.g. Kiowa, one construction in Abun). If there is more than one complementation pattern in these languages, the data points differ with regard to the subordinator, their precise positional behaviour, etc., but their internal structure is fairly similar (leading to a narrow CID range).

The languages that branch out towards the bottom of the network have also been discussed before: Those are the 'strongly desententializing' languages like Yimas, Yukaghir, Evenki, Quechua, Dolakha Newar, Wari', Lavukaleve etc. They are characterized by a comparatively high CID mean (e.g. 7.85 in Yimas, 5.43 in Evenki), and the longer the branch, the narrower the CID range i.e. the less structural variation. Thus for Yimas and Lavukaleve, for instance, it was noted earlier that complementation is rendered exclusively by nominalized constructions, while more sentential types of clause linkage boil down to paratactic direct-speech clauses that have nothing to do with complementation as defined here. If these had been taken into account (e.g. on Noonan's semantic approach to complementation), Yimas and its neighbours would have ended up on the left of the graph, grouping with Mapudungun, Georgian and so on.

A striking property of our network is the fact that most of the sample languages are arranged along the 'path' that ultimately leads to the internally most heterogeneous languages on the left. It appears, therefore, that most languages strike a middle ground in terms of their CID mean and range, so that a sentential type of complement is opposed to one that is significantly decategorized but still clausal enough so as not to qualify as a full-fledged NP. The network is thus the empirically more precise underpinning of previous claims about the structure of complementation systems (e.g. Noonan 2007). However, the fact that no two languages are alike in the middle of the graph but instead create many distinct little edges, testifies to the great diversity that permeates complementation systems across languages. O'Dowd (1992), a study on a much smaller sample and with very different goals from the present one, also finds a certain structural "polarization" of complement clauses, but she suggests measuring "degrees of nominalization" (ibid.: 59) in future research, which is exactly what my study intends to capture. From this perspective, the more the complementation structures in the data deviate from the patterns on the right end of the graph (in whatever direction), the higher the amount of nominalization (somewhere) in the complementation system. For example, when we move away from the extreme cluster on the right and follow a path 'upwards', we encounter (in this order) Slave, Hmong Njua, Burmese, Japanese, Choctaw, Menya, Hup and Lao. These languages have relatively sentential types of complementation overall, but they do involve a significant amount of recategorization in some place of their system. In Hup, one type of

complement has lexical nominalization but a clausal argument structure. In Slave, the internal structure of one complement is fully clausal but it is externally flagged by a clausal nominalizer. In Choctaw, similarly, there is no lexical nominalization partaking in complementation, but some constructions involve conspicuous other structural changes as compared to independent sentences. It is these characteristics that move the complementation systems of Choctaw and Hup (and minimally also Slave) towards the centre of the network, but still locate them in relative vicinity to the 'balanced' complementation systems on the far right.

Degrees of nominalization are arguably most interesting for languages that host several kinds of nominalizing constructions. The multivariate approach taken here is able to capture the structural differences between such similar patterns. Let us consider two cases to illustrate this point. We saw in the network that Dolakha Newar is a language whose major complementation patterns all involve decategorized constructions; the language is thus located towards the bottom of the graph. Two of the complementation patterns involve the same set of nominalizers, but their other properties are sufficiently different, so that it is "best to treat [them] as distinct" (Genetti 2006: 145). The two patterns have markedly different CTP distributions (i.e. they occur in different environments of complementation), which is an effect of their different grammatical properties. Specifically, the first type of nominalization comes without any further subordinator, while the other one contains a clausal nominalizer as a boundary marker. Moreover, whereas the first (unmarked) type of nominalization can only be used with different subjects, the second one can be used in same-subject environments, but then the internal subject NP follows an equi-deletion constraint. Finally, the absence of a boundary marker in the first construction makes it possible for the complement subject to raise into the matrix clause; this is found with perception verbs (cf. the above discussion for raising in such contexts):

(97) Dolakha Newar (Sino-Tibetan, Tibeto-Burman, Bodic: Nepal; Genetti 2006: 142)

Cilā-n   ninpatti āmta   [kho-en   coŋ-gu]   khon-ai.
goat-ERG   daily   3SG.**DAT** cry-PTCP stay-NMLZ see-3SG.PRS
'The goat sees her crying every day.'[54]

Interestingly, in the second type of construction, the obligatory presence of a boundary marker makes this 'argument trespassing' impossible, i.e. Genetti (2006: 146) points out explicitly that no raising can occur, so that in this construction, the complement subject is expressed canonically inside the complement clause (cf. (11) in §2.2.1 for an example). It is these argument-structural and boundary-marking differences that ultimately lead to different degrees of nominalization in Dolakha Newar.

As a further example of such gradation, we can turn to Turkish. It is well-known from the literature that two major complementation patterns in Turkish are based on two different kinds of nominalizers, i.e. *-DIK* and *-mA(K)*, respectively (cf. Csató 2010, Johanson 2011 for their distribution). While the structural properties of the two nominalizations are very similar, there are also some subtle differences between them.

---

[54] In Dolakha Newar, the Dative case is assigned differentially to P arguments that are human and given; other direct objects remain unmarked.

These are insightfully discussed in Kornfilt (1997: 450ff.), who argues that the degree of nominality is somewhat higher in -*mA*-clauses than in -*DIK*-clauses. For example, only the former, but not the latter, have the ability to occur with certain nominal affixes and with demonstratives. It is this property that makes Turkish one of the languages in which there is a slight difference between constructions with one versus several nominal flags: Both nominalizations are case-marked at their boundaries, but the possibility of adding a further boundary marker creates a 0.25-increase in nominality. Interestingly, the occurrence of the relevant demonstratives and other noun markers is not very common overall, and its "acceptability deteriorates with [the] syntactic complexity" (Kornfilt 1997: 451) of the complement: "it is perfect with certain verbs (mostly intransitives)" (ibid.), while transitive and ditransitive predicates appear to avoid it. The ultimate source of this constraint is not discussed, but the description suggests that it is a lexically-specific effect in the first place, which might thus be interesting to explore from a usage-based perspective.[55]

In sum, the idea of the network display presented in Material 5 is that complementation systems can be compared in terms of their degree of desententialization, based on a quantifiable metric of degrees of de- or re-categorization. However, a final problem remains to be tackled in relation to the *NeighborNet* representation of the sample languages. As it stands, the network does not represent the typological similarity of complementation systems entirely accurately. What it plots is the relative similarity of languages according to the CID mean and the CID range, and as I have tried to show, these two parameters provide a revealing comparative picture of the sample languages. But they neglect one important dimension of grammatical organization: Languages that come out as similar on the above calculation may still differ in the *status* that the individual constructions have within their respective system. If we take the two extremes on the left side of our network, Mapudungun and Georgian, one can see that both of their complementation systems involve strongly sentential alongside strongly nominalizing complements. However, a crucial difference between them is the relative importance, or '**systemic relevance**', that the two construction types have in each system. In Georgian, the finite complementation pattern can basically cover all complement-taking environments (except for phasal verbs), while the Masdar (despite being a perfectly regular morphological process) is somewhat more restricted in its applicability and also in its overall frequency (Merab Geguchadze, p.c.). In Mapudungun, the status of the finite/balanced complement is very different because it is restricted to the quotative niche and ungrammatical in any of the other environments distinguished in the present paper. Similar asymmetries are found in many other languages. To give but one more example here, Musqueam (or Downriver Halkomelem) has both nominalized and finite subordinate clauses; in complementation, the finite clauses occur only in clearly demarcated contexts (such as causatives, jussives and indirect questions), while the (clausal) nominalization is productive across almost the entire

---

[55] It was beyond the scope of my study to probe deeper into the Turkological literature, so it may well be that this effect has already been discussed in frequency- or semantically-based terms.

range of CTP environments. It has thus also been described as "the default embedded clause" (Thompson 2012: ii) in Halkomelem, with a much greater systemic relevance to complementation than the finite pattern.

More generally, then, the complementation patterns of a given language are typically ranked according to their versatility (i.e. type frequency) and overall usage prominence (i.e. token frequency).[56] If this dimension is taken into account, the picture of the relative similarities of the sample languages may change in significant ways. In order to find out, I checked the complementation systems for the systemic relevance of each of its associated data points. While a typological study does not have direct access to corpus frequencies in individual languages, I do have a fairly robust picture of the general CTP distribution of each complementation pattern (to be studied in detail in Chapters 6 and 7), which may act as an approximate indicator of the pattern's type frequency relative to the other constructions in the same system. In addition, it is often fairly clear from the description of individual languages which patterns are more versatile or prominent than others, based on the respective authors' original corpora and, more indirectly, on the organization of the description as such. On the basis of these criteria, what I did was the following: If there is only one construction in the system or if there are two patterns without evidence of an asymmetry in their systemic relevance, the mean CID from above remained unchanged. If there are two or more than two complementation patterns in the system and a dominant type could be identified, this type was 'doubled' (i.e. added to the data again, as it were), so that it would have a greater bearing on the mean CID of the respective language. Needless to say, this is somewhat problematic in as far as the CID is measured as a continuous variable but we cannot say (as would strictly speaking be required for continuous variables) that a dominant type is *exactly* twice as relevant to the system as the minor types. However, this is the best approximation of the cross-linguistic situation that I can offer at this point.

In addition to the CID mean, the second parameter of the comparison also needs to be adapted: In the previous calculation, I considered the range that the CIDs span in each language. This does not change if a dominant type is given more weight, but what does change is the standard deviation of the new CID mean. This measure still captures how widely the individual complements of a language differ in their structure, but unlike the range, the standard deviation is calculated on the basis of the CID mean.[57] If the latter is now adjusted to accommodate the weight of certain constructions, the standard deviation changes accordingly. Once these two measures are collected for every language, the procedure is the same as above: a Euclidian distance matrix is

---

[56] In Chapter 3, it was stated that the present research is based on the major complementation patterns from each language, while marginal constructions were disregarded to begin with. But crucially, differences in versatility and usage prominence can, of course, also be found between these central members of each system, and this is exactly what the present section is about. There is thus no contradiction to the selection procedure here.

[57] The mathematical formula of the standard deviation is given below:

$$\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2}$$

created, which can in turn be submitted to a *NeighborNet* computation. The result is displayed in Material 6 (Appendix).

This solution, I believe, captures the structural similarities of the sample languages more faithfully now. We can still observe a major cluster of exclusively nominalizing systems on the left, with increasing structural variation the more one moves towards the centre of the network. (Note that Musqueam, the language discussed above, has now moved a little towards these nominalizing languages, reflecting the higher weight that the new calculation gives to its nominalization construction.) This cluster is still strongly opposed by the non-nominalizing languages on the far right, and many of the relative similarities of the languages in between are structured in analogous ways to the previous graph. A more pronounced change, however, occurs at the top of the graph, where we now see a clearer clustering of languages than previously. The entire 'arm' reaching out from the centre of the network to the top contains languages with relatively high standard deviations, i.e. structurally distinct complements. The further away from the middle, the more differentiated the constructions are (as in the previous network). But there is now a marked bifurcation into languages like Krongo, Ma'di, Turkish, Jarawara and Mapudungun, on the one hand, and languages like Arabic, Jamsay, Tukang Besi and Georgian on the other. The crucial difference between those two branches is precisely the relative weight of the different structures: the complementation systems of Turkish, Jarawara and Mapudungun rely to a large extent on nominalization and only marginally on more balanced structures, while nominalizations in Georgian, Tukang Besi, Jamsay and Gulf Arabic are clearly secondary relative to the finite complements.[58] In other words, the network now contains the split between Mapudungun and Georgian that we intended to capture, and in general reflects the description of the sample languages more realistically (barring the ones for which not all of the relevant structures could be taken into account in the first place).

While further interesting observations could be gleaned from the network display, I will leave it at that for now; more on the organization of complementation systems will be said in Chapter 6 and 7 of the dissertation. My general point in this chapter was to show where the study of complementation structures and the resulting complementation systems can be taken empirically, once a quantitative approach is followed from the local to the global level of grammatical organization.

---

[58] Incidentally, this moves Georgian somewhat closer in the network to the European languages in the sample, such as German, Basque, Hungarian and Serbo-Croatian. This is a welcome change to the extent that subordination in Georgian has been noted to follow European, and specifically even Indo-European, lines: "In Bezug auf die Bildung von Hypotaxen [gleicht das Georgische] ganz und gar einer indoeuropäischen, nicht jedoch einer kaukasischen Sprache" (Bossong 1979: 47).

5

Syntagmatic relations to the matrix

5.1 Introduction

In addition to the morphosyntactic properties of complementation constructions analysed in the preceding chapter, another important dimension of the grammatical structure of complements pertains to their syntagmatic organization. Therefore, the present chapter will be devoted to examining the position of complement clauses relative to their associated matrixes. This issue is part of the larger enterprise of constituent-order typology, pioneered by Greenberg (1963). In this framework, the position of dependent clauses has been studied prominently for relative clauses (e.g. Lehmann 1984, 1986; Dryer 2011b), and recent studies have also made inroads in the domain of adverbial constructions (e.g. Diessel 2001, Schmidtke-Bode 2009, Diessel and Hetterle 2011). For complement clauses, by contrast, little work is currently available that has a sound and comparatively large empirical foundation. To be sure, the position of complement clauses has not gone unnoticed in the literature. Important observations on the behaviour of clausal complements have been made, inter alia, by Grosu and Thompson (1977), Dryer (1980) and Hawkins (1994); all of them have generated an intriguing set of hypotheses on the positional constraints on complement clauses, along with proposals for their possible motivations. These will be discussed in the course of this chapter. However, none of the above studies is based on a sufficiently large and balanced sample of languages, so that, in effect, we have hitherto lacked the relevant data in order to substantiate, refine or refute the claims that have been made. The present chapter attempts to remedy this situation, at least as far as the position of subject complement clauses is concerned. The position of object clauses is studied in the context of the larger research project with which my dissertation is affiliated and will thus be dealt with in separate publications. My primary attention, then, is with subject clauses, thereby adding another layer to our understanding of this type of complement across the world's languages. I will begin, in §5.2 and §5.3, with an examination of the positional patterns of S- and A-clauses, respectively. Although these sections will already contain a fair amount of discussion, especially as far as possible explanatory principles are concerned, it is only in the ensuing §5.4 that all empirical and theoretical strands can be drawn together. The final section 5.5 is devoted to so-

called 'place-holding' elements that can occur when a complement clause is found in a different position from that of the corresponding subject NP. These elements deserve a separate paragraph, not only because they have not been studied in a typological context yet, but because their very existence has important implications for the syntactic status of the complement and the cross-linguistic comparison of complementation systems.

5.2 The positional patterns of S-clauses

5.2.1 Establishing positioning patterns

In this section, we consider the positional behaviour of what we called 'S-clauses' in §2.2. These were defined as complementation patterns that function as the sole argument of a monovalent predicate, regardless of the specific coding that this predicate requires. From a syntagmatic perspective, therefore, the issue boils down to the whether the complement clause preferably precedes or follows the predicate in question, i.e. to the binary contrasts between SV and VS arrangements. By way of introduction, an example of each configuration follows: In (98), we see an S-clause from Ainu in sentence-initial (or preverbal) position, while (99) shows a sentence-final S-clause from Tepehua.

(98) Ainu (isolate: Japan; Tamura 2000: 126)

[*Ku-kemeyki*   *kor*   *kú-itak*   *hi*]   *iyotta*   *pirka.*
 1SG.NOM-sew   while   1SG.NOM-talk   COMP   most   be.good
 'It's best (for me) to talk while I'm sewing.'

(99) Tepehua (Totonacan: Mexico; Kung 2007: 599)

Lhiiyuuch jaantu qox [*nii*   *ka-ta-maqnii-y*   *juu ʔanuʔ luw*].
therefore   NEG   good   COMP   IRR-3PL.SBJ-kill-IPFV   ART that   snake
'Therefore, that they kill that snake is not good.'

In both examples, the S-clause simply follows the regular conventions of constituent order in the respective language, Ainu being a verb-final and Tepehua a predominantly verb-initial language. The position of S-clauses is thus best discussed on the level of languages, where we can compare it to the order of S and V in simple sentences.

    As in any kind of constituent-order research, determining the positioning patterns of phrasal and clausal S-arguments relative to the predicate is not always an easy task to tackle from a cross-linguistic perspective. While the issue is straightforward in languages that have conventionalized constituent orders into syntactic rules, methodological problems tend to crop up when both orders of a certain pattern are attested in a given language. Following an argument by Dryer (1989b), I assume that it is possible to speak of a dominant ordering type if there is a marked preference for it in corpus frequencies (i.e. text counts); Dryer's proposal, which he also applies to all of his word-order chapters in *WALS* (e.g. Dryer 2011c–e), is that a dominant order should be at least twice as frequent as the alternative one (i.e. roughly a 70:30 skewing in performance). For the above-mentioned language Tepehua, for example, frequency counts provided by Kung (2007: 535) reveal that in simple sentences, VS is preferred over SV by a ratio of 80:20. For S-clauses, in turn, we have an explicit statement that

they are even less flexible because they "almost always occur clause-finally" (ibid.: 599). Where no such asymmetry can be established on the basis of explicit statements or frequency data in the description of individual languages, no dominant order can be assigned.[1] It goes without saying, then, that the absence of statements or performance data can make it notoriously hard to classify the language in question. In such cases, I had to go by all examples being provided of the relevant structure, and where these did not yield a clear-cut picture, I again refrained from determining a dominant order.

For the coding of S and V in simple sentences, I was fortunate to be able to check my decision against Dryer's (2011d) database in *WALS*. Where my coding deviated from Dryer's, this is normally because I was able to narrow down the domain of investigation to predicate classes that are relevant for subject clauses. Dryer finds, for instance, that some languages place the subjects of active-intransitive and stative-intransitive clauses in different positions relative to the verb and thus come out as having 'no dominant order' in his coding scheme. However, since I showed in §2.2 that S-clauses normally function as S_U (undergoer) arguments, this issue does not arise in my data to begin with. In other words, where possible (i.e. where lexically-specific information was available), I compared the order of phrasal and clausal subjects only in the *relevant contexts*.[2]

Following the above procedure, then, what I arrived at was a three-way classification of the sample languages:

- Languages which have grammaticalized either SV or VS as a strict syntactic rule, or follow one of these orders as the normal discourse pattern whose exceptions are infrequent and highly marked. For lack of a better term, these languages will be called '**rigid**' SV or VS types. As an example, we may quote from the description of Fehan Tetun, where one-argument clauses are said to "follow a strict SV constituent order" (van Klinken 1999: 13).

- Languages in which both SV and VS orders are attested, but with a frequency bias towards one of the two orders. In contrast to the languages called 'rigid' above, a choice of the alternative order is not considered a highly marked or infrequent pattern. This can be seen when we compare the above description of Tetun to the corresponding one on Chalcatongo Mixtec: "While basic word order in Mixtec is VSO, sentences with one or

---

[1] Note, however, that Dryer himself does not always appear to apply his classificatory criteria consistently. For the sample language Chimalapa Zoque, for instance, he cites Johnson (2000: 378) in support of his classification of the language as SV. Crucially, SV is only preferred for pronominal subjects, while lexical subject NPs, on which his *WALS* feature 82 is explicitly based, show a frequency asymmetry in the opposite direction, i.e. 56:44 in favour of VS (ibid.). Judging by his 70:30 criterion, the language cannot be assigned a dominant value for the order of S and V.

[2] This also applies to languages of the so-called 'Philippine type' (cf. Andrews 2007a: 202 for discussion). In these languages, there are (at least) two kinds of basic clause pattern; our sample language Begak Ida'an conforms to this, having a so-called 'Actor Voice' (AV) construction and an 'Undergoer Voice' (UV) construction existing side by side as canonical clause patterns (i.e. none of them is in any way more marginal than the other, unlike in active/passive systems). Importantly, the two clause patterns are associated with different word orders, both in bivalent and monovalent clauses. In the latter, S-arguments in AV constructions tend to be SV (e.g. 'I go'), while S-arguments in UV are typically VS (e.g. 'sick I', cf. Goudswaard 2005: 126). Crucially, since the predicates that select complement clauses as their S-arguments are all UV-predicates (i.e. stative verbs or adjectives), only the VS order had to be taken as criterial for the comparison between phrasal and clausal S-arguments. Therefore, once the class of phrasal-S arguments is limited to the contexts that are relevant to S-clauses, some languages are no longer mixed SV/VS, but either one of those, so that an ordering pattern can be assigned uncontroversially.

more constituents located in preverbal position are also found. [...] Initial subjects and initial obliques are fairly common; initial objects are relatively rare." (Macaulay 1996: 102) Explicit frequency counts are provided, for example, for Huallaga Quechua (75:25 in favour of SV (Weber 1989: 16)) or Tümpisa Shoshone (3:2 in favour of SV (Dayley 1989: 23)), to name but two. Such ordering patterns will be referred to as '**non-rigid**' SV or VS types.

- Finally, languages in which no dominant order of S and V could be established. This type will be referred to as '**flexible**' here.

In other words, I have split Dryer's binary SV and VS types into rigid and non-rigid patterns, so that four basic positioning types, plus the flexible one, can be distinguished. The coding of S-clauses then followed the same principles. As a result, the positioning types of S-clauses are:

- **rigidly preverbal** or **rigidly postverbal**, respectively; the initial examples in (98) and (99) are representatives of each type.

- **non-rigidly preverbal** or **non-rigidly postverbal**, respectively. Finite complement clauses in German, for example, are more naturally placed in postverbal position but the preverbal one is also possible and not uncommon (Zifonun et al. (1997: 2347) even claim that they might be 'equally unmarked', but corpus studies show clear asymmetries in favour of the postverbal variant, even in the written medium (e.g. Mollica 2010)):

  (100)   German (Indo-European, Germanic: Germany, Austria, Switzerland)

  a. *Es   ist   gut, [dass   du   hier   bist]*.
  3N.SG be.3SG.PRS good COMP 2SG here be.2SG.PRS
  'It is good that you are here.'

  b. *[Dass   du   hier   bist],   ist   gut*.
  COMP 2SG here be.2SG.PRS   be.3SG.PRS good
  'It is good that you are here.'

  The pattern in (100b) is the one that corresponds to the dominant position of S-arguments in simple sentences. S-clauses of this type have thus also been referred to as '**in situ**' in the literature. In (100a), by contrast, the complement is removed from the canonical subject position, this being occupied by the 'anticipatory' or 'dummy' pronoun *es*. Complements of this kind are usually said to be **extraposed**. The topic of extraposition will concern us in more detail later on.

- **flexible**: Both preverbal and postverbal orders of the complement clause are attested and no dominant position could be identified by either the sources or by myself. This holds, for example, for a number of nominalized complements in the data (e.g. Tariana, Georgian (Merab Geguchadze, p.c.) or the Lango Infinitive).

- **adjoined**: This is a special case that we encountered at the very beginning of the dissertation (§2.2); recall that complements may not be embedded in the main clause but adjoined to a full-fledged matrix, resulting in a complementation strategy rather than a complement clause proper. This was illustrated for object clauses, but it is also possible for subject clauses. Although I generally refrained from taking all kinds of paratactic alternatives to subject clauses into account as complementation strategies, two subordinate patterns were taken on board whose status is probably adjoined rather than

embedded. One is basically a correlative clause in Sanuma, which is described as a "subject" (Borgman 1990: 142) but is actually left-adjoined to a matrix that obligatorily contains the third-person pronoun *te*, as an anaphoric copy of the complement clause (as well as an optional relative marker *ĩ*):

(101) Sanuma (Yanomam: Brazil, Venezuela; Borgman 1990: 142)

[*Thomö-mo   sinomo   wi*],   *ĩ*   *te*   *wanisala.*
 steal-PERFORM   habitual   INDEF   **REL**   **3SG**   bad
'Habitually stealing is bad.'

The second case of this pattern is quite parallel. It is found in Menya (Trans-New Guinea, Angan: PNG), where a complement appears on the left periphery of the sentence. In all examples I have of this pattern as an S-clause, it is coreferential with an anaphoric pronoun in the ensuing matrix clause, thus making the subordinate clause a left-adjoined complementation strategy (Whitehead 2004: 209). Note, however, that the same construction can be directly embedded as an internally-headed relative clause (i.e. without an anaphoric element in the matrix), but it is unclear whether this is also possible for its uses as a subject clause. In the absence of evidence of embedding, I have opted for a conservative coding as 'adjoined S-clause'. The two cases discussed here are, in a way, the positional mirror image of the right-extraposed S-clauses above, except that the latter always have an alternative in-situ ordering, while no such evidence is available for the former.

With these ordering types established, we can now proceed to the analysis of their cross-linguistic distribution.

We begin by cross-classifying the ordering patterns of phrasal and clausal S-arguments in the sample (Table 14):

Table 14. Positional types of phrasal and clausal S-arguments and their frequencies

	Pre:rig	Pre:nonrig	Post:rig	Post:nonrig	Flexible	L-adjoined	Total
SV:rigid	48	7	23	10	3	2	93
SV:non-rigid	8	6	2	0	3	0	19
VS:rigid	0	0	14	0	0	0	14
VS:non-rigid	1	0	6	2	0	0	9
Flexible	1	0	1	1	0	0	3
Total	58	13	46	13	6	2	138

Although Table 14 looks confusing at first sight, there are some interesting things to glean from it. To begin with, the total number of complementation patterns ($N_{pat} = 138$) shows that only a severely reduced subset of the overall data for complementation systems ($N_{pat} = 228$), i.e. roughly 61%, has been found in S-function.[3] Secondly, focusing on the basic SV–VS contrasts, without taking rigidity into account, it can be seen that there is a marked difference between the ordering of phrasal and clausal S-arguments, in the following way: Of the 138 complementation patterns in Table 14, 112 (= the totals of the first two rows) come from languages with SV as the

---

[3] Note that two complementation patterns in the data, one from Gooniyandi and one from Modern Khwe, are also known to occur as S-clause, but could not be considered in Table 14 because the data on their positional preferences were not sufficient.

basic order of phrasal S-arguments, while only 23 (= the totals of the 3rd and 4th rows) come from VS-languages. This has nothing to do with S-clauses as such, but reflects in the first place the cross-linguistic dominance of SV over VS: In my sample, 80% of the languages are SV, 16% VS and 4% do not have a dominant order for S and V in simple sentences; Dryer's (2011d) much larger (but non-controlled) sample finds an 86% preference of SV to VS.[4] However, it is now interesting to note that the order of clausal S-arguments does *not* follow this extremely skewed distribution. One can infer from Table 14 that 71 S-clauses (= the totals of the first two columns) are preverbal and 59 (= the totals of the 3rd and 4th columns) are postverbal patterns. This is much more symmetrical than one would expect from the above figures (binomial test, p = 2.1e-14). What happens, then, is that S-clauses in SV-languages commonly deviate from the ordering of phrasal subjects. This is illustrated in the spine plot in Fig. 6 below:
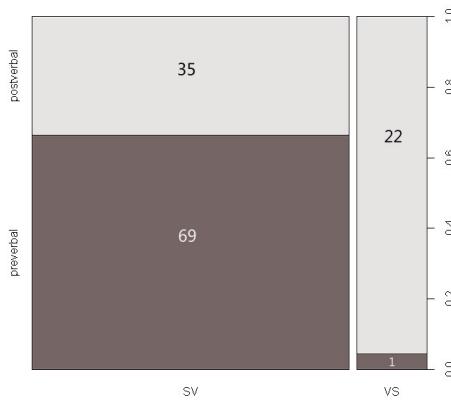
Figure 6. Major positioning types of S-clauses in SV- and VS-languages
(Fisher exact test, p = 2.2e-08, odds ratio = 42.28)

In view of Fig. 6, we can already lend credence to what Dryer (1980) calls the 'final-over-initial' constraint: According to this hypothesis, complement clauses have a greater tendency than phrasal arguments to appear in sentence-final position. The above figures provide statistically sound evidence for this hypothesis in relation to S-clauses. And it can be further corroborated if we take the dimension of rigidity into account. In Table 14, one can see that languages with a rigidly preverbal position of their phrasal S-arguments sometimes loosen this rigidity to allow for some flexibility in their S-clauses (cf. the 7 cases in the second cell of the first row of the table); by contrast, languages with rigidly postverbal order of their phrasal S-arguments generally retain this rigidity for their S-clauses and do not favour the reverse order (cf. the empty 4th cell in the 3rd row of the table). This finding, too, appears to support the 'final-over-initial' constraint. In the following, I will investigate the exact workings of this

---

[4] The motivations for this asymmetry have been discussed under several rubrics in the typological literature. On a discourse-pragmatic approach (e.g. Tomlin 1986), S-arguments are basically grammaticalized topics (cf. Shibatani 1991 for the mechanisms involved) and thus inherit the topical position in the sentence. In processing-oriented explanations (e.g. Keenan 1978), subject-final languages are argued to systematically delay the identification of the referent of whom something is predicated and whose perspective is taken in the utterance, which in turn leads to "a longer 'unprocessed storage' on the part of the hearer" (ibid.: 309). The latter explanation is thus a precursor to Hawkins' (2004) 'Maximize on-line processing' principle, which he applies to a similar ordering asymmetry between A- and P-arguments.

constraint, and the position of S-clauses more generally, by changing to the level of languages: What are the general patterns of grammatical organization that we can observe in regard to the linearization of S-clauses, and how can we explain them?

From this perspective, the sample languages fall into the following categories:

- complete conformity with phrasal S-arguments [N_{lgs} = 40]
- mixed S-clauses [13]
- type shift [11]
- constraining flexibility [13]
- adding flexibility [6]
- insufficient information on the ordering of S-clauses [2]
- no S-clauses attested or allowed [15]

The languages without S-clauses will concern us in more detail in Chapter 6 later on. The areal distribution of the remaining languages for which the position of S-clauses could be determined is shown in Fig. 7 below:
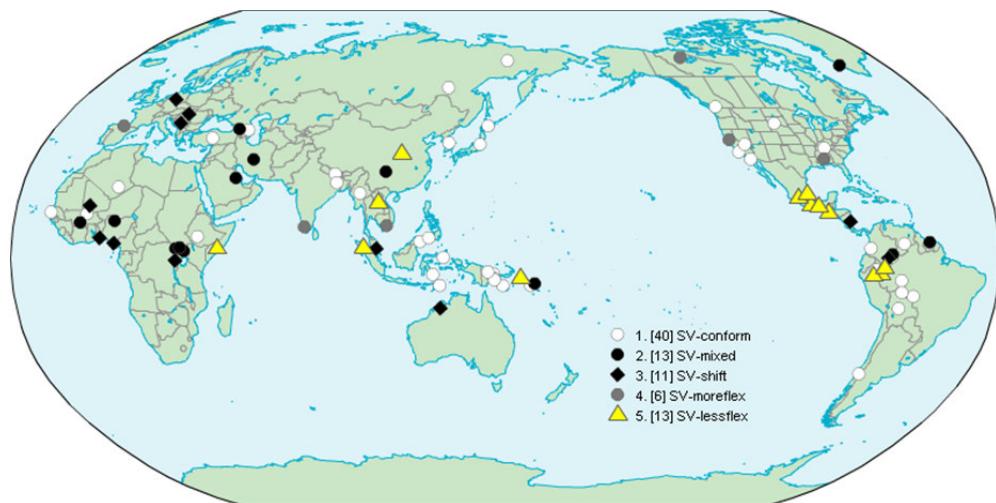
Figure 7. Distribution of the positional organization of S-clauses

In the following sections, each of the types on the map will be discussed in turn. As we go along, various kinds of motivation for the position of S-clauses, and for Dryer's 'final-over-initial' hypothesis, will crop up. These will be collected, as it were, but can be integrated into a more coherent account only at the very end of the chapter, after A-clauses have also been taken into consideration.

5.2.2 The conformity type

The first and largest group of languages is inconspicuous in that it relies on completely analogical behaviour of phrasal and clausal S-arguments. This type is particularly common in my sample around the Pacific Rim (cf. Bickel and Nichols 2006 for this areal pattern), but it also shows up in selected languages in Eurasia (e.g. Lezgian) and Africa (e.g. Jamsay). Structurally, it applies to dozens of SV-languages, but also to VS-languages like Begak-Ida'an (cf. footnote 2 above for why this is classified as VS), Barbareño Chumash, Mapudungun, Musqueam, Tamashek, Tukang Besi, Urarina,

Yakan and Yuracare. In these languages, then, S-clauses generally follow the same linear order as phrasal S-arguments, regardless of the number of distinct S-clauses that are recorded in my database. For example, all four complementation structures in Tamashek follow the verb, in keeping with the VS order of phrasal subjects, while all constructions from Wolaytta precede the predicate, aligning with the rigid SV order of phrasal subjects. For Wolaytta, this *rigid* verb-final ordering pattern is a more far-reaching one, affecting the whole language system. This is illustrated nicely by the following quote:

> "The most important principle on which the Wolaytta syntax is founded consists of the fact that the verbal predicate always concludes the clause it belongs to. Even [when] a main clause governs one or more subordinate clauses, it must always end with its verbal predicate; as a consequence, subordinate clauses are therefore inserted into their main clause. [...] In contrast to some other SOV languages spoken within the Ethiopian cultural area, for instance, Somali and Afar which are characterized by a certain flexibility in the sequence of the single elements within the clause and which thus also allow SVO or OVS constructions, Wolaytta always requires that the verbal predicate conclude the clause." (Lamberti and Sottile 1997: 199)

What languages like Wolaytta demonstrate, then, is a preference of analogical linearization patterns over other factors that may potentially encourage ordering variants. For example, in contrast to English, where the possibility of right-extraposition can be exploited systematically to adapt to the information status of the complement (e.g. discourse-old in-situ complements versus discourse-new extraposed complements, cf. Kaltenböck 2004: 180ff. for a detailed study), no such principled ordering alternatives present themselves in *rigid* SV- or VS-languages. Needless to say, I was unable to study the discourse-pragmatic contexts in which the individual instances of S-clauses in reference grammars occur, so it remains entirely unclear at present how the languages in question take care of different information statuses in relation to S-clauses, or whether such rigid S-clauses only allow of one specific type of discourse status to begin with (e.g. given information). To be sure, we can sometimes find explicit discourse-pragmatic marking on S-clauses even in rigidly ordered constructions. In Awa Pit, for instance, the Infinitive can function as S-clause on evaluative predicates and in this function, it "is often followed by the Topic marker =*na*, as is common for Subjects" (Curnow 1997: 266):

(102) Awa Pit (Barbacoan: Ecuador, Colombia; Curnow 1997: 266)

[*Ashaŋpa=ta   pyan-na*]=na   wat   shi   ki.
woman=ACC   hit-INF=**TOP**   good   NEG   be.NEG.NLOCUT
'Hitting (one's) wife is not good.'

However, as Kaltenböck (2004: 165ff.) argues, two commonly conflated discourse-pragmatic concepts in relation to subject clauses need to be strictly separated: one is the thematic structure in which the subject clause takes part, and the other one its information status in the discourse. From the first perspective, Kaltenböck maintains that subject clauses always constitute **relational topics**, spelling out what the predication of the main clause is about and being commented on by the matrix

predicate. From the second perspective, subject clauses may in principle be more variable, representing given or new information in an ongoing discourse. If they encode relatively familiar information, they are **topics** in a **referential** sense (cf. also Gundel 1988 for this distinction). In English, Kaltenböck (2004: 180) argues, "it is precisely this interaction of information status and topichood that is responsible for a distinctive communicative function of extraposition and non-extraposition." But in the absence of systematic extraposition constructions, it is an open question if and how the information status of S-clauses can be regulated. In Awa Pit, the Topic morpheme primarily indicates *relational* topichood ('aboutness'); whether or not this entails that S-clauses must also represent referentially given information, is thus an independent issue, to which I cannot give an answer at present. More generally, the discourse-pragmatic properties of subject clauses constitute an understudied aspect of complementation cross-linguistically and would hence be an extremely interesting avenue for future research. Obviously, they require a fundamentally different methodological approach and primary materials than the present study. [5]

Similarly to information status, rigid SV- or VS-languages do not appear to have linear means for adapting to the length of the complement clause. As we shall see below, considerations of length (often referred to as 'weight') are important in languages that allow for S-clauses to appear on both sides of the matrix predicate, but again this is not a systematic option for rigid SV-/VS-languages. However, there are indications that this is not at all problematic for the languages in question. If one believes in the validity of a 'short-before-long' principle of syntactic organization (e.g. Behaghel 1909, de Smedt 1994, Dik (1997: 404), Wasow 2002), all rigid VS-languages are unproblematic since the (heavier) S-clause follows the (lighter) CTP. On this account, only SV-languages with in-situ S-clauses would be more challenging to explain. Importantly, my data clearly reveal that such structures are far from uncommon. However, the vast majority of the languages in question (27/30) are head-final languages (i.e. SV+OV in Dryer's (1997) terminology) like Japanese, Amele, Lavukaleve, Newar and so on. Considerations of 'end weight' are unlikely to apply to most of these languages: there is evidence from both corpora and experimental studies that where heavy elements can be moved in strongly head-final languages, the direction of movement is often to the left (e.g. Yamashita and Chang 2001 for Japanese, and cf. Hawkins 2007 for the non-universality of 'end weight'). This is corroborated by some of the other sample languages, such as Amele, where "heavy structures are

---

[5] An interesting finding in Kaltenböck's (2004) study is the medium- and genre-specific distribution of the correlation between extraposition and information status of the complement. In written language, extraposed subject clauses strongly tend to encode new information, while extraposition in spoken language shows a relatively even split between referential givenness and newness (ibid.: 181–83). Moreover, the figures for spoken language vary depending on the specific genre or speech situation: So-called 'unscripted monologues' follow the pattern of written discourse, while 'private spontaneous dialogue' is remarkable because it is the only genre in which the majority of extraposed subject clauses encode given rather than new information. For typological studies, these two categories are particularly interesting since they often constitute the primary genres from which data are drawn (i.e. records of spoken language from either spontaneous settings or from oral narratives, which may well correspond to Kaltenböck's 'unscripted monologues'). His study thus provides very fertile ground for the cross-linguistic study of information status in complement clauses. (Incidentally, his figures also reveal that when S-clauses in spoken private dialogue appear in situ, they almost invariably encode given information. It is thus not unreasonable that this could be a more widespread pattern, bearing ultimately on languages like Awa Pit above.)

normally moved to the sentence initial position" (Roberts 1987: 147). From this perspective, in-situ S-clauses in rigid head-final languages do not pose any alleged production or comprehension difficulties only because of their length (cf. also Hawkins 2004: 110 for production and parsing in head-final languages more generally). If head-final languages are neglected, there are only three cases left (Noon, Taba, Tetun): These are SV+VO in simple sentences but S-clauses adhere to the rigid SV-order of phrasal subjects. We will return to these cases in the course of this chapter, as each of them patterns with certain languages of the other groups.

5.2.3 Partial or full type shifts

In this section, we turn to all languages that are represented by a black spot on the map in Fig. 7, either as a dot or as a rhombus. As can be seen, the bulk of these languages clusters in Eurasia, the Middle East and Africa. However, since it also occasionally extends to languages of very different areas and genetic affiliations, I believe that it cannot be explained entirely in geographical or genealogical terms. What the relevant languages have in common that at least a subset of the complementation patterns in S-function changes its preferred position from SV to VS or vice versa (regardless of the rigidity dimension). We will see shortly that these shifts follow similar principles, irrespective of whether they apply to some or all S-clauses of a given system. For this reason, it is useful to discuss all of the 'black languages' together.

The languages in question are collected in the following table; the groupings will be explained during the exposition.

Table 15. Languages with partial or full type shifts of the position of S-clauses

Partial type shifts			Full type shifts	
A	B	C	SV > VS	VS > SV
Gulf Arabic, Hausa, Hmong Njua, Ma'di, Mayogo, Ndyuka, Persian, Supyire, To'aba'ita	Georgian; West Greenlandic	Lango; Tariana	Fongbe, Kana, Koyra Chiini Nkore-Kiga, Serbo-Croatian Semelai, Ungarinjin German, Hungarian, Rama	Barasano

Languages with partial type shifts are SV-languages and turn out to follow a very similar principle, although it comes in three different subpatterns, labelled A to C in Table 15. The languages in group 'A' all have two or more complementation patterns available in S-function, but there is a categorical structural split: In each case, a 'finite' construction is categorically postverbal, while a 'non-finite' one is kept in situ. This is illustrated for Supyire below:

(103) Supyire (Niger-Congo, Gur: Mali; Carlson 1994: 459)

    a. *La   à   yaa*   [*mìì   í   ŋ́-káré*].
        3SG   PRF be.fitting   1SG   SBJV   INTR-go
        'It's right that I go.'

b. [*U* ̩*ŋ-kàrà-ŋá*]   *à*   *yaa.*
   3SG.POSS NMLZ-go-DEF PRF be.fitting
   'Her going/leaving was appropriate/convenient/a good thing.'

In the literature, patterns such as (103a) are often described as involving 'extraposition'. Noonan (2007: 93), for example, defines extraposition as "the process of moving a complement to the end of the sentence", even where this is an obligatory pattern and not a matter of selecting between positional variants. However, in Supyire (and the other languages in subgroup A), 'extraposition' would be an unfortunate choice of terminology since "there is not a shred of evidence that the sentence-like complements [...] were ever placed in subject or direct object position" and are hence in any way "removed from direct object or subject position and placed outside the clause" (Carlson 1994: 461). Instead, Carlson argues that these finite types of complement clauses have always been found at the right end of the sentence, i.e. right-adjoined to matrix clause. In the case of object complements, the matrix often contains a cataphoric object pronoun, so that the complement itself stands in an appositional relation to that pronoun; and the same holds for S-clauses, which, as (103a) demonstrates, strongly favour the occurrence of an anticipatory pronoun ('it'). While we do not always have historical records for the languages in group A, the synchronic pattern is identical to that in Supyire: the finite S-clause is categorically found in the postverbal position and cannot occur in situ. Where certain leftward 'movements' towards the subject position are possible, they normally have to involve certain structural changes and do not lead to the direct embedding of the finite complement in subject position. In Hausa, for example, sentence-initial position of the "S-clause" is possible, but then the subject position of the matrix clause is still typically filled by either the expletive subject ('it') or a demonstrative ('that'). This so-called "topicalization" (Newman 2000: 105) is a focal operation that also regularly applies to NPs, and it structurally results in a left-adjoined rather than an in-situ subject clause. Other structural changes occur in Persian. This language allows for its *ke*-complement to occur preverbally, but according to almost all sources that I have consulted (including native speakers), the complement then needs to be preceded by the definite article or a similarly nominal 'head' and cannot occur in the same form that it takes postverbally. Compare (104a) and (104b) below:[6]

(104) Persian (Indo-European, Iranian: Iran; Aghaei 2006: 38–39)

    a. [*'In*   [*ke*   *Dāvud*   *varšekaste* *šod-e*]]   *doruq-e.*
      DEM.NOM COMP Davoud bankrupt became-3SG lie-3SG
      'That Davoud has become bankrupt is not true.'

    b. *[*Ke*   *Dāvud*   *varšekaste* *šod-e*]   *doruq-e.*
      COMP Davoud bankrupt became-3SG lie-3SG
      'That Davoud has become bankrupt is not true.'

---

[6] In the formal syntactic literature on Persian (and languages with similar phenomena), this has been formulated as a constraint that a preposed complement clause needs to be embedded in a "DP shell" (Aghaei 2006: 36 and Lotfi 2006, but cf. also Inaba 2007 for comparative notes on this phenomenon in a formal framework). Interestingly, and much to my confusion, the reference grammar by Mahootian (1997: 30) does contain an example of an in-situ finite clause without a demonstrative. As stated above, though, this is judged impossible by all other sources.

In other words, there is a significant structural difference between the 'intraposed' and the allegedly 'extraposed' construction: It is only in this type of 'clausal nominalization' with the demonstrative (or an appositive head noun) that *ke*-clauses can stay in situ. Alternatively, if this position is to be retained, a more strongly nominalized complement, the so-called 'Infinitive' (Mahootian 1997: 144) can be chosen:

(105) Persian (Indo-European, Iranian: Iran; Aghaei 2006: 12)

    [*Dars*   *xānd-an*]   *kar-e*   *ʿāsāni ʿast.*
    lesson   read-**INF**   job-EZ   easy   be.3SG
    'Studying is an easy job to do.'

To summarize so far, the languages in group A exhibit a categorical association between the position and the internal structure of the complement. The languages in groups B and C can be seen as variants of this pattern. In group C, we find Lango, which has a sentential type of complement that must be postverbal as an S-clause, and an Infinitive that freely occurs on both sides of the predicate (no dominant position could be discerned in the data); Tariana is similar (though the syntactic function of the finite clause in question is somewhat unclear; if it can count as an S-clause in our sense, it is a categorically postverbal construction). In group B, Georgian shows a pattern that involves genuine extraposition this time, in the sense that there is variability involved: The sentential type of S-clause can in principle occur before or after the CTP. Its distribution is skewed, though, because right-extraposition is clearly preferred to leaving the complement in situ.[7] The Masdar, by contrast, does not show a similar asymmetry; according to my informant, either order is commonly found and the difference between them may be guided completely by information-structural considerations. (An example of such an extraposed Masdar was given in (36) in §4.2.1.) If this intuition lives up to corpus-linguistic scrutiny, the Georgian Masdar is quite different, for example, from the English Gerund, which, in carefully balanced data, is overwhelmingly preferred in situ (Kaltenböck 2004: 110ff., *pace* Erdmann 1987).[8] At any rate, the pattern is again such that finite and non-finite S-clauses are associated with different ordering preferences. West Greenlandic is similar but not identical: The more sentential type of complement freely appears before or after the CTP; in this language, considerations of end weight play a certain role in the distribution of complements (Fortescue 1984: 95), but there is no strikingly dominant order for finite S-clauses. The nominalization, by contrast, also shares much of this general flexibility, but does seem to be preferred in preverbal position. The generalization underlying Georgian and Greenlandic, then, is that nominalizations are at least as prone to occur in situ as finite S-clauses, thus tying in with the pattern we saw for languages of group A. And what is also in line with this pattern is one of the languages that were

---

[7] Note that, in view of such situations where the postverbal position is dominant, researchers have argued that the label 'extraposition' is actually misleading (e.g. Mair 1990 on English *to*-Infinitives). I fully agree, but I will stick with the term extraposition here in order to refer to the very fact that there is variation between a preverbal and a postverbal ordering pattern.

[8] Therefore, as Kaltenböck (2004: 112) states, right-extraposition of Gerunds "requires a special reason, such as satisfying the principle of end weight". For interesting historical data on the development of the Gerund in subject position, cf. Fanego 2007.

mentioned as 'exceptional' in §5.2.2: Recall that Noon apparently makes its S-clauses follow the rigid SV-order of phrasal subjects, despite the fact that it is not a head-final language (unlike most other data points in the 'conformity group' above). The reason is very simple: The only complementation pattern that can function as subject clause in this language is the Infinitive, the finite complement being categorically ruled out (Soukka 2000: 273), and in keeping with the languages discussed in the present section, the Infinitive is retained in preverbal position. This is why Noon ended up in the 'conformity' group above.

We can now have a look at the right side of Table 15, where we find so-called 'full type shifts'. This basically means that all distinct S-clauses recorded for a given language in the database follow a positional pattern that deviates from that of phrasal S-arguments. As can be seen, this almost invariably involves shifts from SV to VS, and the languages provided in the four rows in this rubric differ only in how exactly the shifts play out. Thus the S-clauses recorded for Fongbe, Kana and Koyra Chiini all appear to follow a relatively rigid postverbal position (no exceptions found), while the corresponding NP subjects would be rigidly preverbal. Note that all of the relevant constructions are structurally of the sentential type of complement.[9] In Nkore-Kiga and Serbo-Croatian, both structural types of S-clause are dominantly postverbal and thus contrast with the SV order of phrasal S-arguments; but while the finite S-clauses is strongly tied to the postverbal position, the non-finite variant appears to leave at least some room for variation. My informant on Serbo-Croatian, for example, largely rejected finite S-clauses in situ, while the Infinitive was felt to be acceptable in this position but is also more naturally extraposed. In Semelai and Ungarinjin, phrasal S-arguments allow for more positional variation than the above languages, but the corresponding S-clause is either stated to be rigidly postverbal (Kruspe 2004: 348 on Semelai) or only attested in this position without any further information (Ungarinjin). In both cases, the complement is of the sentential type. In German, Hungarian and Rama, finally, there are two structural types of complement (finite and non-finite) and *both* types are variable but prefer the postverbal position.[10] By and large, then, the pattern underlying all languages with type shifts from SV to VS adheres to the generalization noted for 'mixed' languages above. Specifically, there appears to be a statistical implication to the effect that if a non-finite S-clause shows the possibility to deviate from the preferred order of phrasal S-arguments, a finite S-clause in the same language will exploit this possibility to at least the same degree.

A type shift of a rather unexpected nature occurs in Barasano, the final language in Table 15. Barasano is an OVS language, with the unmarked position of phrasal subjects

---

[9] For Fongbe and Kana, a non-finite type of complementation exists, but it is of limited system relevance and, more importantly, could not be considered due to a lack of detailed information. In Koyra Chiini, the Infinitive has not been found used in subject functions. There is some potential, then, that these languages ultimately also fall into the group of partial type shifts.

[10] In Hungarian, this seems to be a much stronger tendency than in German, where Infinitives in situ are found more regularly. Depending on the type and amount of data used, the literature on German has come up with conflicting statements as to the positional constraints on the Infinitive as subject (contrast, for example Olszok 1983 (unmarked = 'in situ') with Mollica 2010 (unmarked = 'extraposed')). My own intuitions as a native speaker support the postverbal preference of infinitival S-clauses.

being "sentence-final" (Jones and Jones 1991: 65). Crucially, however, S-clauses appear in sentence-initial position and thus induce a shift from VS to SV:

(106) Barasano (Tucanoan: Colombia; Jones and Jones 1991: 161)

[*Kũbua   kẽdo-re*]   *bɨto   hosa-a-ha.*
 canoe   good.CAUS-NMLZ   very   be.difficult-PRS-3
'Making a canoe is very difficult.'

There are at least two possible explanations of this ordering pattern. On the one hand, the order of phrasal subjects is not rigid but follows a systematic discourse-pragmatic pattern: "The subject is sentence-initial when first introduced or in special focus" (ibid.: 65). In this light, sentence-initial S-clauses may be motivated similarly, but I do not have access to the specific contextual conditions under which the respective S-clauses were produced. On the other hand, however, there is an alternative explanation, drawing on structural analogy. When the nominalized complement is used as an object clause, it always remains in its canonical OVS position, i.e. preverbal or rather sentence-initial. There are thus quite a few tokens of the construction in language use that do not show anything other than sentence-initial position, so this ordering pattern may have become extended to CTPs that take the nominalization as their subject. And interestingly, the analogical pressure is reinforced on a more abstract (i.e. type- rather than token-) level since "subordinate clauses generally precede main clauses" (ibid.: 113).[11] It is unlikely, therefore, that S-clauses end up sentence-initially because of a special information status; they rather appear to be conventionally drawn into this position in keeping with other uses of the same construction and with other subordinate-clause constructions, despite the fact that the S-clause then deviates from the unmarked position of subjects in the language.

The descriptive, and perhaps also explanatory, power of analogy is, in fact, of a much more far-reaching nature, being applicable also to all partial or full type shifts from SV to VS that we have encountered in this section: The S-clauses that deviate from the canonical order of phrasal S-arguments appear in exactly the same position that the relevant complementation pattern occupies as an object clause. This is also noted in Dryer (1980), who explicitly argues for analogy to be at work. The real question, however, is *why* such analogical pressure would exert itself and under which conditions it is relaxed. My answer to these questions is somewhat different from Dryer's, but a full consideration of this point will have to await the inspection of the remaining languages in the sample and also the examination of A-clauses. Therefore, I will first continue the empirical overview of the positioning patterns, before final generalizations and explanatory accounts can be developed.

5.2.4 Adding flexibility

The few and fairly scattered languages coded as grey dots on the map in Fig. 7 add a certain amount of flexibility when it comes to positioning S-arguments, without inducing a more radical shift in ordering preferences. Thus Choctaw, Malayalam,

---

[11] This statement refers specifically to complement and adverbial clauses. Relative clauses in Barasano (which are nominalizations in apposition to a head noun), are more flexible and commonly occur postverbally.

Slave, Vietnamese and Wappo have a fairly rigid SV order for phrasal subjects, but at least one of the available S-clause constructions can also appear postverbally. In Basque, the order of phrasal S-arguments is a little more liberal to begin with, and complement clauses are yet more flexible by being commonly postverbal (although no gross asymmetry could be determined; Trask (1997: 114) suggests the label 'flexible'). At least some of these cases, then, can be referred to as extraposition in the proper sense, and they provide evidence for Dryer's (1980) 'final-over-initial' hypothesis from the arena of language performance. The factors that guide extraposition are normally to be found in either heaviness or information-structural considerations. For Basque, for instance, we know that 'end weight' plays an important role: "As in English, heavy elements, especially clauses, tend to occur to the right in Basque. […] The longer the finite complement is, the more likely it will be for it to be in postverbal position" (Hualde and Ortiz de Urbina 2003: 452, 652).[12] For Malayalam, by contrast, Asher and Kumari (1997: 185) argue that there are "no rules applying specifically […] to 'heavy' structures under which these are optionally or obligatorily moved to some other position in the sentence than the one which they would normally occupy"; complement extraposition thus seems to be motivated by other, possibly information-structural factors (ibid.), although no systematic information is available. For Slave, two things about extraposition are noteworthy: First, although extraposition of S-clauses is possible, "the nonextraposed versions are generally judged as preferable" (Rice 1989: 1196). Extraposition thus seems to be kept in check by the cross-categorial preference for head-final syntax in the language, i.e. by analogical pressures of the language system. Second, and more interestingly, right-extraposition cannot apply to object clauses of the exact same structural type, i.e. those have to stay in situ. Contrast the following examples:

(107) Slave (Na-Dene, Athapaskan: Canada; Rice 1989: 1226, 1004)

a. *Setá   [ʔeyi dene sexé   ʔeghálaundá]   enịdhẹ.*
   1SG.father that man 1SG.with 3.OPT.work.3   wants
   'My dad wants that man to work with him.'

b. *Síré   nezọ   [turi   nadídéh].*
   sure   3.is.good   small.lake   1PL.travel
   'It's really good to travel to small lakes.'

This is a rather interesting distribution from a typological perspective. In keeping with the observation on rigid OV-languages above, object clauses in Slave can be moved to the left periphery of the sentence, following a rule of topicalization that can also

---

[12] Interestingly, in Basque, the heaviness principle may even override information-structural considerations: Normally, the immediately preverbal position in Basque is reserved for focal elements. But in finite complementation, two exceptional constellations may arise. First, the extraposition of object clauses leads to a shift from SOV to SVO structures, thus appearing to leave the subject in the preverbal focus position; however, extraposition is often performed even if the subject is non-focal. This, then, is a tolerated by-product of the desired avoidance of heaviness. Secondly, according to Saltarelli (1988), extraposition of complements is commonly performed even if the complement represents focal information and would thus be expected preverbally. Taken together, these two observations lead one to conclude that heaviness is able to outrank information-structural constraints in certain situations. (Overall, however, the sources on Basque agree that sentential complements are commonly found in both preverbal and postverbal position, and no striking asymmetry can be established.)

regularly be applied to NP objects. Such leftward movement is particularly called for when the object clause is itself more complex (i.e. containing another clause), regardless of whether it represents topical (old, given) information. It rather forestalls "difficulties of processing multiple centrally embedded structures" (Rice 1989: 1206), as predicted by processing theories of centre-embedding more generally (e.g. Kuno 1974, Grosu and Thompson 1977, Hawkins 2004, inter alia; and cf. also §5.3 below). But crucially, this movement is always to the left, while right-extraposition is not an option for object clauses. This is rather reserved for S-clauses and, interestingly, for some adverbial clauses and also relative clauses (cf. Rice 1989: 1194). For the adverbial clauses in question, this is explicitly stated to be due to heaviness (ibid.: 1195), and it appears that the same motivation also underlies the right-extraposition of S-clauses and relative clauses.

In yet other cases, extraposition is not (only) dependent on the syntactic context but a lexically-specific pattern, such that certain CTPs either prefer or even require the S-clause to occur postverbally. In Wappo, for example, the Infinitive construction normally appears in situ as an S-clause, but "with the verbs meaning 'easy' and 'difficult', the Infinitive does not occupy the subject position" (Thompson et al. 2006: 141):

(108) Wappo (Wappo-Yukian: USA; Thompson et al. 2006: 140–41)

    a. [*K'ešu mehlah-ukh*]   *uwa-khiʔ.*
        deer   hunt-INF   bad-STAT
        'Hunting deer is bad.'

    b. *Tahwal'-lah*   [*chica ṭ'ol-ukh*].
        job-NEG.DEP bear   catch-INF
        'It's easy (lit. 'not a job') to catch a bear.'

In Vietnamese, extraposition is possible (though not frequent) for many CTPs of S-clauses; interestingly, however, when the complement predication is to be negated, the S-clause is placed after a negative CTP (typically *không phai (là)* 'not correct (is)'), which itself occurs "at the beginning" of the sentence (Nguyên 1997: 235):

(109) Vietnamese (Austro-Asiatic, Mon-Khmer, Viet-Muong: Vietnam; Nguyên 1997: 235)

    Không phải   (là)   [*tiểu-bang Cali   thiếu tiền*].
    not   correct (be)   state   California lack   money
    'It's not true that the state of California lacks money.'[13]

---

[13] Incidentally, such lexically-specific effects are also found in English, where predicates like *seem*, *appear* or *happen* must extrapose their S-clauses and cannot leave them in situ:

  (110) a. *It seems* [*that Merlin is late*].
             It happened [*that Merlin was late*].

        b. **[That Merlin is late] seems.*
            *[That Merlin was late] happened.

For further discussion of these verbs, cf. Ransom 1986: 95.

The explanation of such CTP-driven effects is likely to reside in the semantics of the predicates as such, but neither the materials on Wappo nor those on Vietnamese offer any further details.

In sum, the languages in the present section preferably leave S-clauses in situ but can relax this constraint under specific circumstances. In the next section, we will illustrate the opposite pattern, i.e. a certain rigidification of the order of S-clauses.

5.2.5 Constraining flexibility

The languages in this group have in common that their S-clauses have been found in fewer positions than phrasal S-arguments. Typically, this means that a language without a dominant order of S and V in simple sentences exhibits an ordering preference for clausal S-arguments (Karo Batak, Matsés), or that a language with some flexibility for phrasal S-arguments does not mirror the same kind of flexibility for S-clauses. The latter group is attested for both SV-languages and VS-languages. In the former, a non-rigid SV pattern appears to be limited to a more rigid one (Lao, Chinese, Motuna, Purépecha, Huallaga Quechua, Somali). In the latter, a non-rigid VS pattern seems to be limited to an exclusively postverbal one, as far as we can tell (Chalcatongo Mixtec, Tepehua, Tzutujil, Yagua, Chimalapa Zoque). As can be seen in Fig. 7, especially the last group has a conspicuous areal profile, comprising a dense cluster of Mesoamerican languages. The distribution of phrasal and clausal S-arguments in these languages ties in perfectly with what Gast (2007) describes as "structural homogeneity despite phylogenetic diversity" in the Mesoamerican area, especially in relation to constituent-order typology. In other words, it is likely that common patterns of linearization in this region are at least reinforced by language contact.

Overall, it has to be conceded that this category is the 'weakest' one, for two reasons. First, the data or statements on the order of phrasal subjects are often of a global kind, attesting to *some* conspicuous flexibility of subjects in certain contexts. It cannot be guaranteed, however, that these are the contexts that are relevant to the assessment of S-clauses, i.e. that they affect the CTPs in question. Thus it was only in certain cases that more 'local' evidence of directly comparable contexts could be assessed. Second, the present group of languages is also the one with the least explicit information on the position of the S-clause as such, so that I had to rely on the (sometimes scarce) data available. Consequently, the class as a whole should be taken with a large grain of salt, as long as we do not know for certain whether the absence of a particular ordering pattern reflects a genuine constraint or merely a gap in the data.

For this reason, I refrain at this point from attempting to provide an explanation that can potentially capture all of the above cases: As long as it is unclear if there is a real constraint here, this is a futile exercise. For the sake of completeness, however, let me mention that the VS-languages above would be in keeping with Dryer's 'final-over-initial' constraint, and for at least some of them, we know that weight effects play a certain role in complementation; this would at least be compatible with a tendency not

to prepose clausal S-arguments.[14] By contrast, Dryer (1980) notes that the languages which rigidify the order of S-clauses to *preverbal* position are potentially problematic for his hypothesis, and hence deserve some comment. For example Mandarin Chinese, in which subject NPs can be postverbal if they encode indefinite information, is rigidly SV for S-clauses (even for those that encode generic and hence 'indefinite' propositions). (Dryer 1980 maintains that the postverbal position is downright ungrammatical, while Li and Thompson 1981 note that it is possible though clearly marginal. In either case, the S-clause does not fulfil the criteria for being called 'non-rigid' in my coding schema.) But this amounts to a situation in which (i) the S-clause is more constrained than the corresponding NPs, (ii) the direction of this effect goes against the 'final-over-initial' hypothesis, (iii) there is no analogical support from the corresponding object clause: S- and P-clauses occur on different sides of the predicate. In this respect, Mandarin Chinese differs from some other SV-languages in the present category, such as Quechua and Motuna; in these languages, the dominantly preverbal position of S-clauses ties in with the head-final syntax of the corresponding object clauses and of other grammatical categories in the language. Chinese is thus truly different here, and so is Purépecha. Regardless of whether a 'rigidification' of the S-clause is an empirically valid assumption for this language, the basic fact remains that the dominant position of its S-clauses is clearly preverbal, while object clauses are postverbal, just as in Chinese. Taken by itself, this combination is perhaps not too unusual, but what makes these languages special is that the in-situ S-clause is not a nominalized, but mostly a sentential structure. This distinguishes them from S-clauses in other SVO-languages: Recall from above that in typical SVO-languages like Gulf Arabic, Hausa, Hmong Njua, Ndyuka or Serbo-Croatian, finite S-clauses are preferably or even categorically constrained to VS position, while the in-situ position necessitates a non-finite, typically nominalized, construction type. In this respect, Chinese and Purépecha are exceptional and hence need some sort of motivation. Languages that we may add here are those that we left open as exceptions in the other categories: In §5.2.2, we noted that Taba and Tetun leave their S-clauses in the same in-situ position as phrasal S-arguments, despite the fact that they are not head-final languages. And in these cases, too, the complement is of a sentential and not a non-finite nature, and the corresponding object clause is postverbal.

Dryer (1980) leaves structural considerations completely aside, and his explanation for the behaviour of S-clauses in Mandarin Chinese is cast in terms of analogy again. Specifically, he claims that while the position of S-clauses differs from that of P-clauses, it is still harmonic with the branching direction of other subordinate clauses in the language, particularly relative clauses. As is well-known, Chinese employs prenominal relative clauses, and this, Dryer claims, provides the necessary analogical support for

---

[14] Certain heaviness effects are attested, for instance, for Tzutujil, where "many [instances of the complement clause in question] are extraposed to the end of the sentence", e.g. object clauses that are moved beyond an indirect object (Dayley 1985: 235). This, then, may also be the reason why S-clauses of the same type are kept in postverbal position, despite the fact that phrasal S-arguments have considerably more freedom to appear preverbally. Note that the preverbal position of subject NPs in Tzutujil is especially common when they represent discourse-old information. However, S-clauses in this language are kept postverbally even where they arguably encode given information. This suggests an explanation in terms of heaviness even more strongly.

keeping subject clauses in situ. He interprets this multiple analogy of subordinate clauses of different types as falling out from an overarching architectural principle of language that eases syntactic processing: "An underlying assumption of syntax is that constituents tend to share the positional properties of constituents to which they are syntactically similar." (Dryer 1980: 174) While I am generally sympathetic to explanations in terms of analogy, the specific explanatory power accredited to relative clauses in Dryer (1980) does not extend to the other languages above: Purépecha, Taba and Tetun have postnominal relatives. It is thus questionable that these cases follow an overarching principle to align subordinate clauses for ease of processing. On a more local level, however, it is true that at least Purépecha and Taba (and Chinese, for that matter!) have a significant amount of word-order freedom as compared to typical SVO languages, including the possibility of OV order in transitive clauses. From this perspective, then, 'left-branching' or preposing is not an unusual constellation for them, which may indeed have some facilitative effect to loosen the positional inertia that characterizes finite complements in rigid SVO languages. We will return to this point in the more general discussion in §5.4 below. Before we can do so, however, the position of A-clauses needs brief consideration, so as to complete the empirical picture.

5.3 The positional patterns of A-clauses

In many ways, the positional patterns of A-clauses are very similar to those of S-clauses above, so that we can keep this section rather short; our focus will be on the basic distributional facts and on interesting language-specific differences to S-clauses. The major complication that arises in relation to A-arguments is the added layer of syntactic complexity in transitive clauses[15]: A-arguments normally occur with verbal rather than adjectival predicates, and there is another argument, the P-phrase coding the experiencer, involved. As a result, A-clauses can theoretically occur in any of the permutations that A, V and P allow in transitive clauses, i.e. AVP, APV, VAP, VPA, PVA, PAV, or without a dominant order to begin with. And these patterns can, of course, be further specified with regard to their rigidity. In order to keep the distinctions manageable and meaningful, the following overview breaks the data down into coarser-grained categories. I distinguish between A-initial (AVP/APV), A-medial (VAP/PAV) and A-final (VPA/PVA) patterns in simple sentences. These occur with the following positional types of A-clauses in my data (Table 16):

Table 16. Positional types of phrasal and clausal A-arguments and their frequencies

	Initial:rig	Initial:nonrig	Final:rig	Final:nonrig	Flexible	Total
A-initial	30	3	7	4	3	47
A-medial	0	0	2	0	0	2
A-final	0	0	3	2	0	5
Total	30	3	12	6	3	54

---

[15] It is relatively safe to use the term 'transitive' here since the present section will focus on A-clauses that we called 'canonical' in Chapter 2. Most of the so-called 'non-canonical' A-clauses, whose syntactic status is often less clearly that of an A-argument, are not taken into account so as not to distort the picture.

Again, there are several interesting things to be gleaned from this overview. First, the total number of constructions has shrunk down to about one fourth of the overall number of data points, comprising merely 54 complementation patterns (from 37 languages) that can be used in A-function (and whose position in this function could be determined with at least a certain degree of confidence). Second, constructions from A-initial languages (i.e. the first row of the table) outnumber all others by far; this is again due to the fact that A-initial orders in simple sentences are way more common across the world's languages than A-medial or A-final structures: In my sample, 80% of the languages are A-initial, 7% A-medial, 8% A-final and 5% do not have a dominant position for A-arguments in simple sentences; Dryer's (2011c) much larger but again non-controlled sample exhibits a similar (80.3%) preference for clause-initial A-arguments.[16] However, just as with S-clauses above, the position of A-clauses is not as skewed as that of the corresponding A-NPs in simple sentences: In Table 16, we can see that 33 sentence-initial A-clauses (the sum of columns 1 and 2) oppose 21 non-initial ones (the sum of columns 3, 4 and 5); this distribution differs significantly from the expected one (binomial test, $p = 0.002$). Again, the major contributing factor here is the deviant patterning of some complements in A-initial languages (Fig. 8):
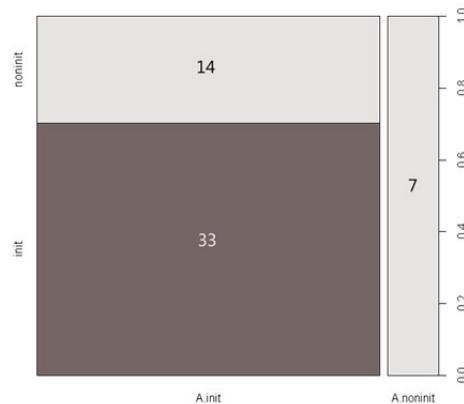
Figure 8. Major positioning types of A-clauses (y-axis) in A-initial- and A-non-initial-languages (x-axis) (Fisher exact test, $p < 0.001$, odds ratio = inf.)

What we are observing, then, is a considerable amount of type shifts from A-initial to A-final, which basically affect the same languages as above (e.g. Gulf Arabic, Choctaw, German, Hausa, Hmong Njua, Hungarian, Persian, Serbo-Croatian) and in much the same way: A sentential type of complement is either fixed or significantly preferred in sentence-final position, and if there is a non-finite complement in addition, it either gravitates more markedly to the sentence-initial position or simply shows the same sentence-final preference as the finite type. A nice example of such a 'finite' construction comes from Choctaw, which is A-initial in simple sentences, but "with psychological verbs that take *-cha* and *-na* clauses as complements, the preferred position is after the verb. Sentences with psychological verbs in which the *-cha* and *-na*

---

[16] This figure was calculated by taking the 1,053 A-initial languages in his sample as the basis, to which I added those 53 languages which have two dominant orders in transitive clauses and where one of those is A-initial. This yields a total of 1,106 out of 1,377 languages which are commonly A-initial in simple sentences.

clauses precede the verb are also acceptable, but rarely volunteered." (Broadwell 2006: 275) The dominant position is illustrated in (111) below:

(111) Choctaw (Muskogean: USA; Broadwell 2006: 274)

Sa-nokoowa-chi-h   [*Pam-at*   *hamburger*   *a-hokóopa-na*].
1SG.II-angry-CAUS-TNS   Pam-NOM   hamburger   1SG.III-steal:LGR-COMP:DS
'It made me mad that Pam stole my hamburger.'[17]

What would also fall under the rubric of a type shift is the sentence-final position of A-clauses in languages that leave phrasal A-arguments clause-medially, i.e. in VAP or PAV position. Interestingly, there is only one such A-medial language in my sample for which A-clauses are attested at all (Tamashek).[18] In this language, all relevant instances of A-clauses available to me (finite or nominalized) occur sentence-finally, so this seems to be indicative of a type shift; however, we cannot be entirely sure since the object argument in such cases is always pronominal, and such pronominal objects appear as person indexes on the verb rather than NP arguments. Therefore, the critical contexts to decide on a potential type shift (i.e. those involving a lexical object NP) are missing. An example containing a pronominal object follows:

(112) Tamashek (Afro-Asiatic, Berber: Mali; Heath 2005: 678)

I-kk`irœḍ=tœn   [*a=d*   *ə̂jlə-n*].
3SG.M.SBJ-obligate.RES=3PL.M.OBJ   DEM=COM   go.SHRT:IPFV-3PL.M.SBJ
'It obligates them that they go.' (i.e. 'They must go.')

While it is thus possible that a lexical object argument would occupy its canonical position after the complement, this would be exceptional from a cross-linguistic point of view: It has been argued in various places in the typological literature that (finite) complement clauses in sentence-medial position strongly tend to be avoided if they contain a clause-initial subordinator (Grosu and Thompson 1977), which is precisely the case in (112) above. Dryer (1980) went on to show that the avoidance of a centre-embedded complement clause holds more generally, i.e. it has also been found (albeit more weakly) for complements with a final or without a subordinator. This led him to formulate the 'initial-over-internal' and the 'final-over-internal' hypotheses, respectively: According to these two hypotheses, complement clauses eschew centre-embedded positions more often than the corresponding NPs, moving more freely to either the sentence-initial position or to the sentence-final position. In both Grosu and Thompson (1977) and Dryer (1980), these tendencies are explained by recourse to sentence processing: The basic idea is that a centre-embedded clausal structure, by virtue of its syntactic complexity, interrupts the parsing of the main clause and hence

---

[17] Argument-structure constructions are extremely involved in Choctaw, with a complex set of partially 'conflicting' rules for case assignment and indexation. While some verbs (e.g. 'bother') do appear to take P-clauses rather than A-clauses as their complements (unlike English), the present example is arguably an instance of an A-clause: The predicate 'angry' is causativized, normally takes a Nominative-marked A-argument (but finite complement clauses in A-function are not case-marked), a class-I (= zero) A-index, and a class-II P-index (*sa-*). Therefore, the argument structure seems to be such that the complement clause is really in A-function.

[18] In the other sample languages with a preference for A-medial structures in simple sentences, I have not been able to find A-clauses in the materials available to me. These languages include Krongo, Chalcatongo Mixtec, Musqueam, Warao, Yagua and Yakan.

delays the construction of "the main sentoid relations" (Grosu and Thompson 1977: 144). Moving the complement clause to a marginal position in the sentence thus normally allows for a more *efficient* way of constructing these relations, as has been argued in several publications by John Hawkins (e.g. Hawkins 1994, 2004). Hawkins (2004: 139) also applies this line of argumentation to VSO languages with A-clauses, which he claims to "regularly extrapose a centre-embedded clausal subject" (although no data are provided). This claim is complementary to Dryer's (1980: 32) inability to find "any case of a VOS language which allows the VOS word order when the object is sentential" (cf. also Keenan 1978 for similar data). From this perspective, then, the sentence-final position of A-clauses in Tamashek (a VSO or rather VAP language) would not be unusual at all. It is thus likely to represent a type shift as well.

Overall, then, 9 of 37 languages with A-clauses in my sample have been found to exhibit a partial or total type shift in the position of A-clauses, with most of them replicating the structure–position interaction of their S-clauses. The remaining languages also fall into the groups we encountered for S-clauses above. This is shown in Fig. 9:
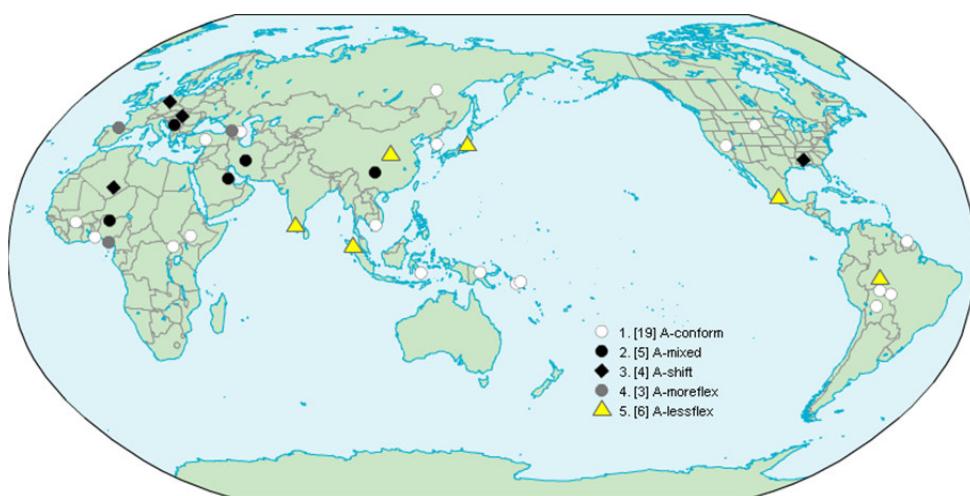
Figure 9. Distribution of the positional organization of A-clauses

As can be seen, we have to recognize a rather large 'conformity type' again (N_{lgs} = 19), where the available evidence does not indicate a significant positional difference between phrasal and clausal A-arguments. Interestingly, however, the membership of this group is now somewhat different as compared to S-clauses above. While it still (expectedly) includes strongly SV/OV-languages like Amele, Evenki, Lezgian, Korean, Kwazá, Lakota, Lavukaleve, Turkish and Wolaytta on the one hand, and VS/VO-languages like Tukang Besi, Wari' and Yuracaré on the other, it now also features a number of languages that belonged to a different group in regard to their S-clauses. These 'new members' of the conformity group, notably Fongbe, Ma'di, Ndyuka, Supyire and To'aba'ita, follow the same principle, though: They all have multiple complementation patterns, comprising more sentential and less sentential (typically nominalized) structures; as S-clauses, the sentential construction is postverbal while the less sentential one occurs in situ (thus the languages came out as 'S-mixed' above).

Crucially, while both construction types are available for S-clauses, only the less sentential type has been found to work as an A-clause. And since it is typically nominalized, the A-clause can remain in the same in-situ position as phrasal A-arguments, creating a 'conformity type'. The example from Supyire provided in (113) completes the pattern we saw for this language in (103) above:

(113) Supyire (Niger-Congo, Gur: Mali; Carlson 1994: 458)

    [U   *ǹ-jà-ǹ-jìrì-ɲa*]   *a*   *mìì* *páà.*

    his NMLZ-be.able-NMLZ-get.up-DEF   PRF me surprise

    'His being able to get up surprises me.'

In effect, then, these languages follow the same coding principle as the 'A-mixed' group above, but have radicalized the pattern in that they do not allow the sentential complement in A-function to begin with. They thus provide further evidence for a correlation between the structure and the position of subject clauses. Note, finally, that Vietnamese is also a new member of the conformity group, but for a different reason. As we saw above, S-clauses in Vietnamese can be extraposed beyond the CTP (and even need to occur postverbally with negative predicates). According to my sources, which are confirmed by my informant, A-clauses are less prone to be extraposed, i.e. they strongly tend to remain in situ. (Extraposition is not ungrammatical, but apparently very infrequent, fulfilling the conditions for my 'rigid' positioning type.) As a consequence, S-clauses in Vietnamese allow for some positional flexibilization, while A-clauses largely conform to the position of phrasal A-arguments.

    We do, however, also find the opposite pattern, i.e. a flexibilization of A-clauses as opposed to phrasal A-arguments. The three languages in question are marked by the grey dots in Fig. 9. In Basque, we find exactly the same tendency for A- as for S-clauses: In both cases, the nominalized complement is very flexible in its position, and known to be amenable to right-extraposition (cf. the discussion above; interestingly, the finite type of complement cannot be used as an A-clause at all). In Georgian, too, the Masdar is flexible in its position (as we noted earlier) and hence found at the left sentence boundary less often than the corresponding phrasal subjects. This appears to be true for Masdars in both S- and A-function. Interestingly, however, there is also a difference between S- and A-clauses, namely with regard to the finite construction type. We saw above that finite S-clauses are favoured in postverbal position, but, according to my informant, finite A-clauses are preferred sentence-initially: while extraposition is possible, it is stylistically marked (unless the complement occurs in a 'DP shell' with a nominal antecedent like 'the fact'). The unmarked position would thus be the following:

(114) Georgian (Kartvelian: Georgia; Hewitt 1995: 613)

    [(*I+s* *gare+mo+eba,*)   *rom*   *e+s* *xalx-i*   *a+s+e* *advil-ad* *t'q'u-i-s*],

    that circumstance.NOM COMP this people-NOM so   easy-ADV lie-PRS-it

    m-a-cvipr-eb-d-a.

    me-NV-amaze-TH-IMPF-it

    '(The fact) that this people so easily resorts to lying used to amaze me.'

Another case of flexibilization, and a simultaneous contrast to S-clauses, is found in Kana. This is a rather rigid SV/AVP language, where phrasal subjects normally occur preverbally. Complement clauses in S-function are categorically postverbal, which is why they were included in the 'type shift' category for S-clauses above. A-clauses, by contrast, occur in either order with respect to the predicate, i.e. they can be left in situ but also be extraposed. Therefore, this pattern constitutes a flexibilization of A-clauses as compared to both phrasal A-arguments and clausal S-arguments:

(115) Kana (Niger-Congo, Benue-Congo, Cross River: Nigeria; Ikoro 1996: 281–83)

    a. *Lēē*   [*kɔ̀ɔ̀*   *ḿ-lū*   *lɛ̀ɛ̀*   *wāá-kīī*].
        good   COMP   1SG.DEF.FUT-come   before   you.PRFV.FUT-go
        'It is good that I should come before you leave.'

    b. *(Á)-wēè yè-rè*   *sàŋ*   *būù-nàdū*   [*kɔ̀ɔ̀*   *ŋwíí*   *gbò*   *tó*   *tó*].
         it-PST   insert-CAUS   anger   stomach.ASSOC-Nadu   COMP   child   Gbo   cry.FACT   cry
        'It angered Nadu that Gbo's child cried.'

    c. [*Kɔ̀ɔ̀*   *ŋwíí-gbò*   *tó*   *tó*]   *wēè yè-rè*   *sàŋ*   *būù-nàdū.*
        COMP child.ASSOC-Gbo   cry.FACT   cry   PST   enter-CAUS   anger   stomach.ASSOC-Nadu
        'Nadu became angry that Gbo's child cried.

Apparently, then, A-clauses allow for in-situ positions while the corresponding S-clauses do not. This is quite a remarkable pattern, so it will be worth returning to in the general discussion later on.

The final category on the map captures rigidification, i.e. cases in which the positional flexibility of A-arguments is constrained when these are clausal structures. The corresponding 6 languages are marked by yellow triangles in Fig. 9. As with S-clauses, the category as a whole is rather weak since it is based on limited evidence (only very detailed corpus studies could corroborate the validity of the group). Within these limits, it appears that Japanese, Jarawara and Malayalam follow a similar pattern: They are basically verb-final languages (Japanese and Jarawara more strongly so than Malayalam), but the relative order of A and P in simple transitive clauses is variable: While A is preferred in sentence-initial position (yielding APV), the alternative object-initial order (PAV) is also possible. With A-clauses, however, this flexibility appears to be constrained at least statistically: The A-initial order is overwhelmingly preferred, so that PAV structures are more exceptional than in simple sentences. One possible reason for this is again the 'initial-over-internal' preference: In the APV order, the A-clause would appear at a sentence margin, which is beneficial because it usually outranks the P-argument in length. This order is thus again more processing-efficient than a centre-embedded PAV structure, and it would also be consonant with the shift of heavy elements to the left in at least Japanese.[19] A slightly different line of reasoning may apply to Mandarin Chinese. As is well-known, Chinese is a topic-prominent

---

[19] In Malayalam, there is a lexically-specific exception to this preference: the verb 'seem' takes a subject clause as stimulus and a Dative-marked NP as experiencer (cf. §2.2.1 and especially example (14) for discussion of this pattern). In this construction, the experiencer is usually topical, so that the complement clause is centre-embedded between the experiencer NP and the matrix verb. However, this is only an A-clause under the assumption that both the complement and the experiencer NP are obligatory constituents; more typically, 'seem' is a CTP for S-clauses.

language (Li and Thompson 1976), so that the clause-initial position of a phrasal A-argument is commonly overridden by having a topic phrase precede the A-argument:

(116) Mandarin Chinese (Sino-Tibetan, Sinitic: China; Li and Thompson 1981: 15)

Zhāngsān wǒ yǐjīng jiàn-guo le.
Zhangsan I already see-EXP CR
'Zhangsan, I've already seen (him).'

When the sentence contains an A-clause, however, it seems much less common that this is preceded by a topic phrase, thus leaving the A-clause in sentence-initial position. An example was given in (34) earlier on, and another one is provided below:

(117) Mandarin Chinese (Sino-Tibetan, Sinitic: China; Li and Thompson 1981: 480)

[*Wǒ cóng mén-kǒu chū-lai*] *bǎ Wángèr xià-le yi-tiào.*
1SG from door-mouth exit-PRFV OBJ Wanger frighten-PRFV one-jump
'My coming through the door gave Wanger a scare.'

Complements like in (117) may be seen as instantiating the A-argument of the transitive main clause (with the NP *Wángèr* being overtly marked for its direct-object status). On an alternative interpretation, however, such A-clauses may themselves be the topic of the sentence, with the grammatical subject of the main clause ('it', 'that') being null-instantiated (Chinese does not have 'dummy' or 'expletive' subjects). Li and Thompson (1981: 480) describe these cases in such a way that "the understood subject [of the main clause] refers to a proposition." Therefore, while the grammatical status of these complement clauses is indeterminate, their discourse-pragmatic status as relational topics ('starting points') seems relatively uncontroversial. This, in turn, may explain why they are not typically preceded by yet another topic phrase (e.g. 'Wanger, [I came through the door] scared]'), at least less commonly than phrasal A-arguments.

A final interesting candidate for a possible rigidification of A-clauses is a complement clause from Karo Batak. In this language, AVP is the unmarked order in simple transitive clauses, but verb-initial VPA and VAP structures are also possible (cf. Woollams 1996: 186–189). Crucially, A-clauses have been attested in the author's corpus only if the object is also expressed by a complement clause, and then they follow the canonical AVP order:

(118) Karo Batak (Austronesian, Western-Malayo Polynesian, Sundic: Indonesia; Woollams 1996: 310)

Arah dahlin bagé, [*mulihen kerina temué*] *njadiken* [*ngadi kerja mengket*
in way thus go.home all guests ACT.cause stop feast enter

rumah pengulu Ajinembah].
house headman Ajinembah

'In this way, all the guests returned home, bringing about an end to the house-warming feast held by the headman of Ajinembah.'

It is unclear whether these 'double-complement' structures obey a genuine constraint, but if they truly constitute the only possible environment for the occurrence of A-clauses, then their positional properties may be explicable in the following way: In the order given above, the two complement clauses appear on either side of the predicate,

so that their syntactic functions can be conveyed clearly. Changing to the VAP or VPA pattern would result in having two complement clauses in immediate succession. This by itself is probably an unusual constellation. It is aggravated by the fact that complements in Karo Batak normally go unmarked, and even if an optional complementizer were inserted to indicate the boundaries between the two clauses, there are no further signals to cue their correct syntactic functions: Both VAP and VPA are possible candidates for alternative orders, and there is no morphological marking on the complement clauses that would flag them as being the subject or object of the main clause. From this perspective, sticking to the unmarked AVP order makes perfect sense: The two complement clauses are nicely separated by the matrix verb, and they appear in the most reliable positions for subjects and objects.

This case concludes our survey of the positional preferences of A-clauses. In this and the preceding section, we have observed a number of recurrent patterns for the linear organization of S- and A-clauses, and a host of possible motivations that may explain these patterns. In the next section, I will attempt to weave these pieces into a coherent whole that captures the language-specific and cross-linguistic observations alike.

5.4 Discussion

When the linearization of S- and A-clauses is investigated from the perspective of linguistic systems, the dominant organizational pattern that we saw was the so-called 'conformity type'. The relevant languages are characterized by the fact that they leave their S- or A-clauses (or both of them) in the same position relative to the predicate as the corresponding phrasal arguments. In head-final (SV+OV) languages, this yields preverbal, and typically sentence-initial, subject clauses. We saw that some of these languages have a tendency to move heavy elements to the left sentence margin, or are more likely to place an A-clause sentence-initially (preceding both the object and the predicate) than embedding it in the centre of the main clause. The latter tendency was taken as evidence for Dryer's 'initial-over-internal' hypothesis, and I agree with the proposals which seek to explain these ordering patterns by recourse to efficient sentence processing. The mirror image of these tendencies is found in head-initial (VS+VO) languages, which yield postverbal, and typically sentence-final, subject clauses. There was some evidence for Dryer's 'final-over-initial' hypothesis in these languages, in as far as S-clauses showed a more rigid adherence to the postverbal position than phrasal S-arguments (e.g. Tepehua, Tzutujil), and also possibly for the 'final-over-internal' hypothesis in relation to A-clauses (recall our discussion of Tamashek (and of object clauses in VPA languages) in this connection).

Cases like A-clauses in Tamashek lead us naturally to the second general pattern in the sample, viz. the significant proportion of the data that is characterized by at least partial deviance from the preferred order of phrasal subjects. These cases almost invariably worked in the direction from SV/AV to VS/VA, involving either a positional flexibilization as compared to phrasal subjects (e.g. SV:rigid → SV:non-rigid) or a genuine shift in the preferred ordering pattern (e.g. SV:rigid → VS:non-rigid/rigid). It is these phenomena, then, that corroborate Dryer's 'final-over-initial' hypothesis most

strongly. In some instances (19/138 S-clauses and 9/54 A-clauses), the positional departures are characterized by their synchronic variability, i.e. by the fact that they represent the outcome of an ordering choice. In these cases, a postverbal subject clause is commonly chosen in language performance, but the alternative order would also have been a grammatically licensed option. One may hence be justified to speak of 'extraposed' subject clauses, and we saw that the reasons for extraposition can be manifold. A factor that is often discussed explicitly is that of syntactic weight or 'heaviness'. We saw this at work for some SV+OV languages like Basque, West Greenlandic or Slave, in which (some types of) subject clauses can occur postverbally, and especially so if they are long. Corpus-linguistic studies, particularly on English (Erdmann 1987, Mair 1990, Kaltenböck 2004), have demonstrated that extraposition often depends more specifically on the length *differential* between the matrix clause and the subject clause, and not primarily on the absolute length of the complement as such. Erdmann (1987: 43) refers to this as the 'principle of complementary weight', which has also been emphasized repeatedly in Hawkins' (1994, 2004) work on constituent order. If this line of argumentation is correct, it may harbour some potential for explaining differences in extraposition between S- and A-clauses: All three of the above corpus studies find that adjectival predicates and intransitive verbs in the matrix clause have a significantly greater likelihood of extraposing their subject clauses than more elaborate VPs, and the latter crucially also include transitive matrix clauses that contain an A-clause as their subject. Therefore, the following constellation is typical:

(119) English (Kaltenböck 2004: 220, 258)[20]

    a.   *It's very sad* [*to see how many are going out of business at present*].

    b.   [*To keep a criminal in jail*] *costs 300 pounds a week.*

Although there is no perfect correlation between the in-situ position of A-clauses and the extraposed position of S-clauses, a certain weight difference between their matrix clauses can often be established. In this light, the positional differences of finite S- and A-clauses in Georgian and Kana would make sense: Recall from above that finite complements in Georgian tend to be postverbal as S-clauses but preferably in-situ as A-clauses; in Kana, only A-clauses have the potential to occur in the canonical subject position (compare (115a) and (115c) again). Therefore, the fact that, in both languages, A-clauses are relatively more open to the sentence-initial position may be related to the inherently more complex nature of the matrix clause of A-clauses: In comparison to S-clauses, the environment of A-clauses creates a smaller length difference on average between the matrix clause and the complement clause, which is why the latter are more open to non-extraposed variants.[21]

---

[20] The examples have been chosen in such a way that they both include the same structural type of subject clause (e.g. the Infinitive) and occur under comparable discourse-pragmatic conditions (in both cases, the complement clause encodes given information relative to its preceding discourse).

[21] In this connection, a few observations by Erdmann (1987) are insightful: First, he shows that extraposition in English is obligatory for intransitive modal predicates (*That he is right may be*) and certain other intransitive verbs if they are not extended by any further material (e.g. *emerge, follow, transpire*). In these cases, then, the weight differential between the S-clause and the matrix clause is particularly great, and in-situ positions become

However, while this is an appealing line of inquiry, it is also possible that positional differences between finite S- and A-clauses are determined by semantic and discourse-pragmatic factors. Kaltenböck (2004: 282) observes, for example, that "matrix predicates expressing a cause-effect relationship between the subject clause and its predicate, such as *to cause*, tend to resist extraposition as this would reverse the logical order" of events. In other words, this is the place where iconicity of sequence may be involved in the linearization of subject clauses (cf., inter alia, Haiman 1985 and Diessel 2008 for the phenomenon in other types of clause combination), leading to the in-situ position of the relevant A-clause.[22] Furthermore, many of the adjectival and intransitive predicates taking S-clauses are "less dynamic" (Mair 1990: 27) than the verbal predicates taking A-clauses: they are commonly stative predicates, including a sizeable number of predicates that "have a fairly low degree of communicative dynamism, i.e. contribute little to the further development of the communication, and are typically only weakly stressed. Their main function is that of asserting or 'staging' the new information in the complement clause" (Kaltenböck 2004: 239). These predicates include, for instance, *it is interesting (to note), worth (noting), important (to remember), happen, turn out, seem*, etc. From this perspective, these discourse-pragmatically "less informative" CTPs "are more suited for non-final position" (ibid.: 281). Overall, it is thus likely that considerations of relative weight are not the only factor influencing the decision to right-extrapose an S-clause or to keep an A-clause in situ. What is uncontroversial, however, is that a number of subject clauses in especially SV+OV languages are principally open to right-extraposition in performance, thereby contributing to the 'final-over-initial' bias in the sample.

This situation contrasts with one in which a subject clause is fixed in a position that deviates from that of phrasal subjects, where 'fixed' is taken to mean that it does not normally occupy the in-situ position without any structural changes being involved. Due to this positional inertia, I argued above that the label 'extraposition' is infelicitous here. The relevant languages are dominantly of the SV/VO type (i.e. AVP in transitive clauses) and cluster densely in Africa and Eurasia (e.g. Hausa, Gulf Arabic, Lango, Mayogo, Fongbe, etc.); however, the pattern has also been found elsewhere (e.g. Hmong Njua, To'aba'ita and Ndyuka). The existence of the pattern has, of course, not gone unnoticed. Dryer (1980: 143), for example, mentions that "many SVO languages exhibit a tendency for sentential subjects to occur in clause-final position; however, it is often not clear whether there is a rule of extraposition involved, and often there is reason to say that there is not." As was discussed above, Dryer's account of these languages rests on the principle of analogy: He notes that, in all of the relevant SVO

---

increasingly possible if more material is added to the matrix clause. Second, Erdmann also shows that A-clauses are not uniform in their positional constraints: Transitive verbs that normally occur with a short (often pronominal) object (e.g. *It strikes/bothers/worries me* [*that …*]) are preferred in extraposed position because the matrix is shorter on average than the ensuing A-clause; by contrast, transitive verbs that often involve longer and more complex objects (e.g. *entail, lead to, make, prove*) prefer in-situ A-clauses. Therefore, the overarching principle uniting these observations appears to be a consideration of 'complementary weight', indeed.

[22] A language supporting this line of argumentation is Yuracaré. In this language, A-clauses are commonly sentence-final (just like phrasal A-arguments), but predicates like 'cause' typically take sentence-initial A-clauses ('Playing football made me sweat', cf. van Gijn 2006), leading to an iconic sequence of events.

languages in his data, the postverbal subject clause occurs in the same position as object clauses and also, with the exception of Chinese, that of relative clauses. Recall from above that Dryer's ultimate explanation for this analogical ordering relies on an overarching architectural pressure to ease the processing of complex sentences: The more consistently 'right-branching' or 'left-branching' the organization of different types of complex sentences, the more the parser can come to rely on a single "perceptual strategy" (Dryer 1980: 173) for the processing of complex sentences. If object clauses and relative clauses are right-branching while sentential subjects are left-branching, Dryer argues, their deviant behaviour "would then upset [this] perceptual strategy."

While intuitively appealing, this line of explanation becomes less convincing when seen in broader dimensions. For one thing, I have already shown that a number of languages in my sample do not exhibit a cross-domain harmonization of their subordinate clauses: In some languages, object clauses and relative clauses are right-branching while subject clauses (of the same structural type) are left-branching (recall Purépecha, Taba, etc. from above, to which we will thus have to return later on). In other languages, subject and object clauses are left-branching structures but relative clauses are not. We discussed the interesting case of Barasano above, in which subject clauses deviate from the subject-final nature of the language to show up in the same sentence-initial position as object clauses. This is also the dominant position of adverbial clauses in the language, and so I argued that analogy may be at work, indeed. However, relative clauses are exempt from this, because they freely occur before and after their head noun in Barasano (and are apparently more common postnominally). Given that the position of relative clauses is crucial to Dryer's argument, Barasano is thus also a potential counterexample to his processing account. Apart from these potential empirical problems, there is also a more general argument as to why Dryer's explanation for analogy may be met with scepticism. Recent developments in linguistic typology and usage-based theory have cast doubt on the idea that the analogical organization of different linguistic subsystems is motivated by holistic processing pressures. For example, Hawkins' (1994, 2004) overarching processing account of cross-categorial harmony in constituent order is opposed by number of alternative explanations that rest in local, i.e. pairwise or chained, analogies that are directly connected by historical processes (Justeson and Stephens 1990, Aristar 1991, Bybee 2010, among others). The present study is not the place to lay this out in more detail, but the general impression conveyed by such alternative proposals is "that global explanations such as Early Immediate Constituents and the Branching Direction Theory [of word-order correlations] may not be necessary, and that the proper generalizations require explanations in terms of local interactions between pairs of word orders" (Croft 2003: 75). These developments and epistemological considerations may also cast a certain amount of doubt on Dryer's proposed architectural principle for the parsing of complex sentences.

With this in mind, we can now return to the languages at stake, namely those (mostly) SV/VO-languages that keep finite subject clauses in sentence-final position. In all cases, it is not only significant that the object clause occurs in the same postverbal

position but, more importantly, that it is of the exact same morphosyntactic type. In other words, we are dealing with the one and the same structure. This is vital because Dryer's (1980) proposal assumes structural similarities between different types of subordinate clauses only on more general level (e.g. that they have the internal structure of a clause (cf. Dryer (1980: 174)). What we are observing here, by contrast, is morphosyntactic *identity*, and this is indicative of the fact that the subject and object clause in question are directly related diachronically. In fact, it turns out that the historical pathways in the above languages are fairly transparent, and they suggest that the complementation pattern first arose as a postverbal object clause, from which it was analogically extended to CTPs that take a complement as subject. The following examples will serve as representatives of recurrent diachronic sources involved in these processes:

(120) Hmong Njua (Hmong-Mien: China; Harriehausen 1990: 238)

Meej kawg [*has tas Peter qaug cawv*].
obvious   COMP Peter drunk
'It is obvious that Peter is drunk.'

(121) Fongbe (Niger-Congo, Kwa: Benin; Lefebvre and Brousseau 2002: 116)

É nyɔ̀ [*nú à ní yì*].
3SG be.good COMP 2SG SBJV leave
'It is good that you leave.'

(122) To'aba'ita (Austronesian, Eastern-Malayo-Polynesian, Oceanic; Lichtenberk 2008: 986)

'E abu la'u bo'o [*na 'o sore-'e*
3SG.NFUT be.forbidden ADD ASS   COMP 2SG.NFUT say-DETR

'oki lae ma=i sa-na].
2SG.FUT go VENT=LOC GL-3SG.PERS
'It is also forbidden for you to say you will come to her.'

One diachronic pathway that clearly involves the rise of object clauses prior to subject clauses is exemplified by (120) from Hmong Njua, where the complex complementizer *has tas* transparently derives from two quotative verbs ('say, speak', cf. also Jarkey 2006: 120 for the same information on White Hmong). The emergence of complement clauses from constructions of direct reported discourse has been documented extensively (e.g. Lord 1993, Güldemann 2008), and as (120) demonstrates, such complements can also be analogically extended to subject contexts. From the relevant set of languages under discussion here, this quotative pathway is also attested for one construction in Hausa, Kana, Fongbe and most likely Mayogo[23].

Example (121) from Fongbe involves the marker *nú*, which is a benefactive preposition and purposive conjunction. This suggests, in keeping with a cross-linguistically common trajectory (e.g. Haspelmath 1989), that the complement is derived from a purpose clause. Importantly, the purpose clause is a postposed

---

[23] In Mayogo, the complementizer *me* is not a verb of saying, but probably a manner adverb 'so, like', which surfaces at the end of the matrix clause in direct speech ('He said so/thus, "[DIRECT QUOTE]"'). As such, it perfectly ties in with Güldemann's (2008) findings for African languages more generally, which commonly have a 'quotative index' that involves an element encoding manner (rather than a quotative predicator).

construction, and the complement clause (whether subject or object) retains the position of the diachronic source. For the languages in question here, this diachronic scenario is also very likely to apply to finite constructions from Ndyuka, Nkore-Kiga and Serbo-Croatian[24].

Example (122) from To'aba'ita illustrates a complement clause that is synchronically identical to a relative-clause construction: *na* regularly surfaces as a subordinator in headed relative clauses, headless relative clauses, appositive clauses and as presupposition marker in focus ('cleft') constructions (similar to English); its use in complementation is more sporadic, but it can occur in both subject and object complements. It is thus at least possible that the complementation pattern in *na* is historically derived from the corresponding relative clause. A similar argument could be constructed for finite *kaa*-complements in Koyra Chiini (cf. Heath 1999: 282f. for discussion). While the relevant historical data are missing on the two languages, the pathway from postnominal relative clauses to postverbal complement clauses has been well-established for other languages in which subject clauses also exhibit a type shift from SV to VS. Prominent cases include German (*dass*-clauses), Hungarian (*hogy*-clauses), Georgian (*rom*-clauses) and Persian (*ke*-clauses), all of which have been argued in the recent literature to be diachronic offshoots of relative clauses in appositive contexts (e.g. Axel 2009 on German, Hendery 2012 on Hungarian and Georgian, Öhl and Korn 2006 on Persian). The basic idea of this diachronic scenario can be illustrated schematically as follows (adapted from Bayer's (2001) discussion of complements in Bengali):

(123) a. *Boy <u>the story</u> knows [*$_{REL}$*that his father him ____ told].*   [relative clause]

  b. *Boy <u>the story</u> knows [*$_{REL}$*that his father come-will].*   [appositive clause]

  c. *Boy _____ knows [*$_{REL}$*that his father come-will].*   [deleted 'head' noun]

  ↓ REANALYSIS

  d. *Boy knows   [*$_{COMP}$*that his father come-will].*   [complement clause]

While in these cases (e.g. German, Persian), the complement clause most likely developed from a right-extraposed relative clause, this is not necessary in SV/VO languages, where relative clauses are right-branching structures to begin with. When complement clauses in such languages thus emerge via this relative/appositive channel, they simply inherit the postverbal position of the object-modifying relative clause.[25]

These three historical trajectories can potentially account for the great majority of the languages in question, i.e. SV/VO languages with subject clauses in rather fixed postverbal position. If they are empirically valid, it follows that no overarching

---

[24] Concerning Serbo-Croatian, the complementizer *da* has been reconstructed for Proto-Slavic, and its origin may be found in the imperative form of the verb 'give'. In Old Church Slavonic (representing the earliest written South Slavic available), *da* was used to introduced purpose clauses and optatives, but was crucially distinct from the subordinator *jako* introducing complement clauses (Wayles Browne, p.c.).

[25] In all cases for which this pathway is described, prototypical matrix verbs include 'know', 'understand', 'see', etc. with an object NP ('that', 'the fact', 'the story') to which the appositive clause relates. The reanalysis is thus likely to first result in object complement clauses. This is supported indirectly by languages in which this pathway has yielded object complement clauses, but no subject complements (e.g. *khā*-clauses in Dolakha Newar, which are "restricted to the O [function]" in the main clause and cannot be used as subject clauses (Genetti 2006: 141)).

processing drive for the analogical ordering of subject clauses, object clauses (and partly also relative clauses) needs to be resorted to; processes of reanalysis and analogical extension, whose ubiquity has been established beyond doubt (cf., e.g., Bybee 2010), yield those parallels of linearization 'for free'. On this account, one may say that the primary reason why subject clauses occur on the same postverbal side as object clauses is that the former are diachronic extensions of the latter by a process known as lexical diffusion (cf. Wang 1969), i.e. the transfer of a complement clause to different semantic and syntactic types of CTPs. However, the arguably more interesting question is now why, in this diffusion process, subject clauses are not drawn into the position canonically associated with subjects in the relevant languages, i.e. the preverbal one. It is on this plane that cognitive motivations may become relevant, after all. Let us consider some possible cognitively oriented proposals in turn.

Maintaining a historical line of inquiry, it may, for instance, be argued that the positional inertia of finite subject clauses is due to entrenchment (Langacker 2000: 3): As we will see in Chapter 6, object clauses typically outrank subject clauses by far in terms of both their type frequency (i.e. the number of different CTPs they attach to) and their token frequency in language use. From this perspective, the pattern 'CTP + finite postverbal complement' does not only represent the diachronic starting point of the construction, but is also continuously reinforced in language use. The production of subject clauses may thus simply be drawn into this powerful attractor template and hence retain a parallel position. This is, of course, entirely speculative, but it resonates well with usage-based accounts of other diachronic phenomena: It has often been observed that frequently activated and hence highly entrenched linguistic patterns resist structural changes (e.g. analogical levelling in inflectional paradigms (Bybee 1985) or accommodation with system-wide syntactic rules (e.g. Bybee and Thompson 1997 on English auxiliaries and their diachronically conservative syntactic behaviour as compared to main verbs)). In a similar manner, finite subject clauses that have been created and entrenched as postverbal constructions may thus resist positional changes suggested by the system at large.

Another factor that may enter the equation is that, conspicuously, all of the examples (120) to (122) above contain a clause-initial subordinator. It has been argued in the literature that the tendency towards sentence-final position of subject clauses is greatest in the presence of such a clause-initial morpheme (Grosu and Thompson 1977, Dryer 1980). The explanations that have been offered for this phenomenon, while all relating in some sense to language processing, are quite heterogeneous. Klaiman (1976) and Dryer (1980) agree that subordinators are functionally most helpful if they occur in between the matrix and the complement clause, so as to clearly signal the boundaries of two clauses (on the more general assumption that clauses are the fundamental units of language processing). Sentence-initial subject clauses with an initial complementizer ([*That he dislikes Mary*] *is obvious*) cannot serve this function and are hence selected against (cf. Dryer 1980: 168). Other approaches (e.g. Grosu and Thompson 1977, Hawkins 1994, 2004) refer to the inefficiency in processing the 'main sentoid relations' (as we saw above) when the sentence begins with a subordinator: In Hawkins' framework, the subordinator is a processing help in that it signals right from

the start that a complex sentence ultimately needs to be constructed, but, given this early information, the presence of the ensuing subject clause then delays the construction of the matrix clause elements considerably.[26] Leaving the subject clause in postverbal position is thus again the more efficient arrangement, primarily for the weight effects discussed above: All immediate constituents can be recognized and attached in a relatively small processing window if the matrix clause is shorter than the complement, which is typically the case in finite subject clauses (cf. *It is obvious* [*that he dislikes Mary*]; but recall the discussion of potential differences between A- and S-clauses in this regard). In essence, then, this type of processing account is but a special case of the rightward-oriented weight effects in SV/VO languages more generally. In light of this, the following two-step scenario may be plausible: With respect to the emergence of the postverbal complement clause, the presence of a clause-initial complementizer is primarily a persistence effect of grammaticalization (in Hopper's (1991) sense): the original quotative, purposive and relative clauses already had the relevant marker, and it occurred right in front of them. It is thus no surprise that, in SV/VO languages, postverbal complements derived from these sources also end up having clause-initial complementizers. The reason why the complement clause remains postverbal even when it has come to function as a subject, may then be due to weight effects, where these are known to be operative for the language in question. In Koyra Chiini and Gulf Arabic, for example, "there is a tendency for any heavy structure to be moved to the end of the sentence where possible" (Holes 1990: 110 on Gulf Arabic); this is why finite object clauses in Arabic are moved towards the ultimate right margin of the sentence (e.g. beyond an indirect object), and it may also be a motivation for leaving finite subject clauses postverbally. An optimal distribution of weight may thus play an important role in determining the position of subject clauses.

However, there is yet more. For one thing, we also know of SV/VO languages with a rigid postverbal position of subject clauses for which weight effects can be ruled out. A case in point is Nkore-Kiga, where "heavy subjects are normally retained in subject position" (Taylor 1985: 79), and yet the finite complement clause in *ngu/ku* remains postverbal even as a subject clause:

(124) Nkore-Kiga (Niger-Congo, Benue-Congo, Bantoid: Uganda; Taylor 1977: 252)

   Nikimanywa [*ngu*   *Mukama waitu akakomooka omuri Yuda*].
   is.evident   COMP   Lord   our   came   out.of Judah
   'It is evident that our Lord came out of Judah.'

In other words, weight is not universally applicable to the SV/VO languages in question, so the position of the subject clause is likely to be motivated differently. Apart from entrenchment, one other fact about many of the languages with postverbal finite

---

[26] Note that Grosu and Thompson (1977) also attempt to engineer a superficially similar processing account (which is why they mentioned together with Hawkins here), but as Dryer (1980) argues in detail, their specific proposal for the allegedly low acceptability of sentence-initial subject clauses is related only very indirectly to the kind of online processing considerations that are at the core of Hawkins' theory. According to Dryer, their account is logically flawed and based on rather dubious assumptions that hearers are supposed to make in 'accepting' or 'rejecting' incoming sentences in discourse. Hence, according to him (Dryer 1980: 165), "their explanation fails to account for the positional tendencies of sentential NPs."

subject clauses is worth recalling here, namely that a non-finite or nominalized construction is perfectly possible in situ. This, too, seems like a straightforward length effect at first sight: 'Non-finiteness' or 'nominalization' suggests structural compression, and hence a considerable reduction in length as opposed to a finite subject clause. But this is not necessarily true. Kaltenböck's (2004: 264) corpus findings on English, for instance, demonstrate that while in-situ Gerunds are shorter on average than their matrix clause (5.9 > 7.8 words, although no significance test is provided), this does not hold for in-situ Infinitives, whose average length outstrips that of their ensuing matrix clause considerably (10.9 > 7.1 words). Moreover, when we think of languages like Persian, the finite *ke*-complement becomes perfectly acceptable in sentence-initial position when it is embedded in a nominal construction, i.e. preceded by an appropriate head noun or article (cf. example (104) again). Without this nominal support, the construction becomes ungrammatical in situ. Given that the nominal in-situ variant is not at all more compact than the postverbal *ke*-clause, considerations of length or complexity cannot explain this. It appears instead that the preverbal subject position requires expression formats that are either visibly desententialized in their internal structure (e.g. non-finite verb forms) or, perhaps more importantly, 'recategorized' in the sense of Chapter 4 (i.e. made nominal in appearance, sometimes without internal morphosyntactic compression). In order to test this intuition, we can now draw on the coding procedure developed in the previous chapter. What we would expect is that preverbal subject clauses might show a higher proportion of dependent verb forms, on the one hand, and a higher average index of desententialization, on the other.

We already know that the SV/VO languages under consideration here exhibit an often categorical association between the verb form and the position. But if we test this against *the entire sample* of subject clauses, it emerges as a statistical tendency of more general relevance. As can be seen in Fig. 10, preverbal subject clauses have a significantly higher proportion of dependent verb forms than postverbal subject clauses (in both S- and A-function). For S-clauses, this difference between initial and final position in regard to the verb form is significant at $p < 0.0001$ (Fisher exact test), and for the whole distribution (including flexible S-clauses) at $p < 0.0001$ (randomized Chi-squared test with B = 100,000, $\chi^2 = 30.45$). For A-clauses, the difference between initial and final position in regard to the verb form is significant at $p = 0.018$ (Fisher exact test), and for the whole distribution (including flexible A-clauses) at $p = 0.03$ (randomized Chi-squared test with B = 100,000, $\chi^2 = 6.41$).
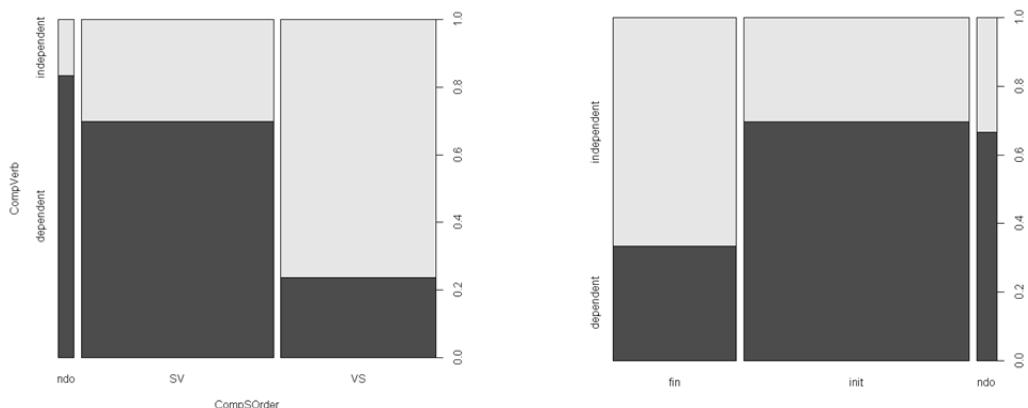
Figure 10. Distribution of dependent and independent verb forms over S-clauses in initial (SV), final (VS) and flexible (ndo = no dominant order) position (LEFT GRAPH), and over A-clauses in sentence-initial (init), sentence-final (fin) and flexible (ndo) position (RIGHT GRAPH)

Apparently, then, lexical nominalization and related processes are relatively more important when it comes to coding propositions in preverbal position than in postverbal position. A similar picture can be obtained if we compare the numerical degree of desententialization for pre- and postverbal subject clauses in the sample. Recall from Chapter 4 that each complementation pattern in the data was allocated a cumulative index of desententialization (CID), which can now be compared for the different positional types of subject clause.[27] This is now a completely different type of variable, i.e. continuous rather than categorical, so an appropriate test for comparing the means of two populations has to be chosen. As was outlined in §3.1, we have to make use of non-parametric testing procedures, and this boils down to treating the data as ordinal, comparing the medians rather than the means of the two distributions (cf. Gries 2013: 215, 230 for discussion). Fig. 11 below plots the CID properties of pre- and postverbal subject clauses. A Wilcoxon test substantiates what the two graphs indicate: The median CID for preverbal S-clauses (**3.5**, *IQR* = 3.5) and postverbal S-clauses (**0.5**, *IQR* = 1.68) is significantly different (W = 3183.5, $p < 0.0001$).[28] In a similar way, the median CID for sentence-initial A-clauses (**3.35**, *IQR* = 4.0) is much higher than that of sentence-final A-clauses (**0.5**, *IQR* = 4.5), although only in a marginally significant way (W = 199, p = 0.05). Note, however, that the weaker signal of A-clauses as compared to S-clauses is due to the fact that A-clauses generally show a slightly higher propensity for nominal constructions, even in sentence-final position; they are thus even more restrictive than S-clauses in regard to structural properties (cf. also Chapter 6).

---

[27] It was pointed out in Chapter 4 that the CIDs calculated there had to disregard some variables of subject clauses since these were not applicable to the entire sample. In the present calculation, by contrast, I only worked with the subsets of the data that *are* used as either S- or A-clause. Consequently, the relevant variables were taken on board in the CID calculation.

[28] *IQR* denotes the so-called 'interquartile range', which is often provided as an appropriate measure of dispersion in the context of ordinal data (cf. Gries 2013: 230).
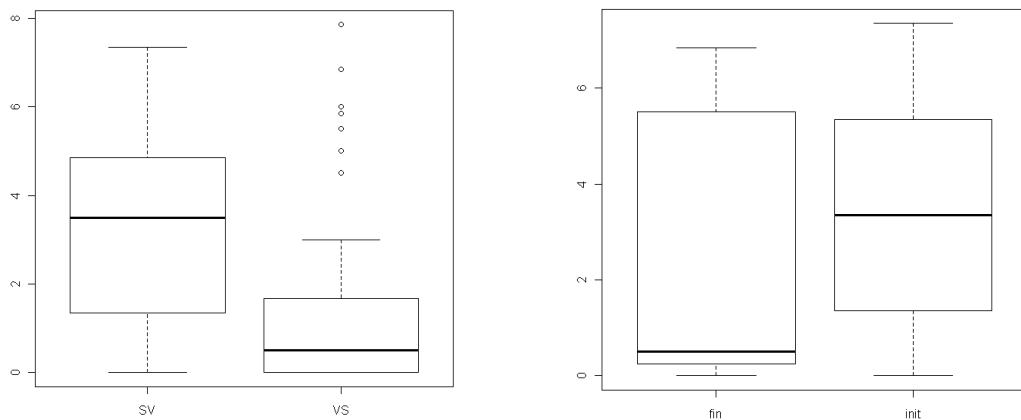
Figure 11. Comparison of the median CIDs in pre- and postverbal S-clauses [LEFT GRAPH], and in sentence-final versus sentence-initial A-clauses [RIGHT GRAPH]

We can conclude from these statistics that there seems to be a general interaction between the position of subject clauses and their structural make-up, one that extends from SV/VO languages to SV/OV types as well. To give but one example: Awa Pit is a rigid SOV language and, in accordance with that, it has a tendency to keep its subordinate clauses in preverbal position as well. Importantly, though, the preverbal types are nominalizations (Infinitive) or participles/converbs, while a sentential type of complement typically follows the matrix clause. Since S-clauses can only be rendered by the Infinitive (while the other complementation patterns are reserved for different CTP classes), Awa Pit contributes a data point in which S-clauses, along with the other preverbal types of complement, are desententialized. In fact, cross-linguistic research suggests that this pattern may extend to subordinate clauses more generally. Lehmann (1984) observed that prenominal relative clauses often take the form of participial (or otherwise clearly deranked and compact) structures, while postnominal relative clauses much more freely allow for the subordinate clause to be a finite, or balanced, construction (cf. also Wu 2010). For the object complement clauses in present sample, Schmidtke-Bode and Diessel (subm.) obtain similar results.[29] Therefore, one may argue that there is a common pressure on left-branching subordinate clauses, not necessarily to be short or 'light', but to be overtly desententialized. For complement clauses, more specifically, this means that they need to show at least some characteristics of the NPs they replace if they also wish to appear in the positions that are normally reserved for them. If they do not meet these requirements, they are kept 'elsewhere'.[30] For subject clauses, this can mean that they remain in the postverbal position in which they historically originated (e.g. in the SV/VO languages considered here, or in SV/OV languages like Supyire, as seen above), or that they can be extraposed from the

---

[29] The larger project with which this dissertation is affiliated is the first to examine the interaction between the structure and the position of subordinate clauses systematically across different clause types. For this reason, I will not go into the relevant findings and issues any further here.

[30] This principle has also been argued, by Güldemann (2008: 239), to be at work in constructions which serve the expression of 'indirect reported discourse' but also some non-reportative complement functions in addition: Particularly in verb-final languages, "indirectly reported clauses display, much more often than directly reported ones, a grammatical treatment that is similar to that of nominal constituents […]. [Importantly,] by making [them] more similar to a nominal entity, the preverbal object position becomes more appropriate to [them]."

preverbal position in performance (e.g. in other SV/OV languages like Choctaw). The general conclusion from this analysis is thus that certain positional slots in the sentence are associated, at least statistically, with requirements on the morphosyntactic structure of subordinate clauses, and that this may be an important factor that motivates the positional inertia of subject clauses that historically arose in postverbal position.

The complex set of motivations adduced in the present section (grammaticalization, entrenchment, weight and its processing implications, structural requirements), especially but not exclusively in relation to SV/VO languages, can be summed up in a slogan that Mair (1990: 27) applies to his corpus data on English: "Avoid using a clausal subject, but if you do, make it a gerund or else try to extrapose it." What I have tried to show in the discussion is, however, that care needs to be taken in employing the term 'extraposition': Especially where a finite subject clause is synchronically fixed in postverbal position, no evidence is available that it has in any way been removed from the canonical (preverbal) subject position; instead, diachronic considerations often suggest that the complement arose in this position (from source constructions that themselves had been right-branching for good reasons). And even for languages like English, German or Georgian, where there is synchronic variability in the ordering of finite S-clauses, the diachronic trajectory and the strong synchronic skewing of the complement to postverbal position suggests that 'extraposition' is the wrong label. The corpus-linguistic studies cited repeatedly in this chapter (e.g. Mair (1990: 30) and Kaltenböck (2004: 279)) are unanimous in that what really needs to be motivated, especially for *finite* subject clauses, is '**intraposition**' (cf. Emonds 1972 for this term).

In this spirit, the remaining part of the present section will be devoted to a closer look at the conditions which favour the intraposition of structurally balanced subject clauses, i.e. we turn to those languages again for which this is actually attested. In English, the in-situ position of subject clauses in general is strongly associated with an anaphoric function, resuming information from the previous discourse in order to make it available for a comment (or 'appraisal', as we said earlier); the ensuing matrix predicate is hence in focus (cf., e.g. P. Miller 2001 and Kaltenböck 2004 for detailed corpus-linguistic evidence of this claim). As Erdmann (1987: 50) puts it, sentence-initial subject clauses thus typically function as a 'retrospective signal' in a text; they are thus commonly found at the beginning of a new paragraph, but never at the ultimate beginning of a new text. This "cohesive function" (Kaltenböck 2004: 278) is reminiscent of so-called 'tail-head linkage' and 'summary-head linkage' commonly discussed for paragraph-initial adverbial clauses (e.g. Longacre 1968). An explicit argumentation to this effect from a more 'exotic' language (albeit one that is not part of the present sample as such) is found for Kokota. Here, Palmer (1999: 294) notes that "the use of a main clause with a subordinate clause as subject occurs frequently in exposition as *a recapping device* indicating the progression of events" (emphasis mine, KSB), as in the following example:

(125) Kokota (Austronesian, Eastern-Malayo Polynesian, Oceanic: Solomon Islands; Palmer 1999: 294)

O   la roh-i   ia   guanha ...
2.SBJ go scrape-TR the.SG guanha
'You go and scrape the guanha [tree] …'

[[*La roh-i   n-e*]   *nhigo-u*],   *toke-na   fa   blahi.*
  go scrape-TR RL-3.SBJ be.finished-PROG   arrive-that.N CAUS be.tabu
'**Going and scraping** finished, [then] go back and bless [it].'

While the sentence-initial position of subject clauses is the norm and not an exception in Kokota, the same discourse-pragmatic motivation probably applies to SV/VO languages with intraposition (though detailed discourse-pragmatic studies such as those for English would be necessary). And from this perspective, it also makes perfect sense that, for the languages discussed above, only nominalized constructions are kept in situ: Not only are these formally closest to NP subjects (and thus meet the structural requirements for the subject position), they also commonly tie in with Givón's (1983) principle of 'topic continuity': According to this principle, discourse-familiar and hence referentially topical material strives towards sentence-initial position and towards economical coding: the more continuous and hence accessible a referent in discourse, the less material its coding requires. The most overtly nominalized constructions, such as gerunds, are indeed often significantly shorter than sentential complements[31] in intraposed position and hence comply with Givón's observation.

Returning to the discourse pragmatics of *finite* subject clauses in intraposition, a final interesting observation should not go unmentioned: Kaltenböck (2004: 254) shows that while intraposed subject clauses as a whole strongly tend towards the textual function just outlined, *that*-clauses in intraposition actually deviate somewhat from this general skewing: in Kaltenböck's corpus, a sizeable number of intraposed *that*-clauses actually encode textually new rather than given information. On the one hand, this pertains to subject clauses whose proposition is referentially new, but which contain an explicit "anchor" to the previous discourse:

(126) English (Kaltenböck 2004: 276)

Implicit in the court's reasoning is the assumption that ownership is absolute or it's not ownership. [*That this is wrong*] *hardly I think needs demonstration.*

As can be seen, the proposition that 'something is wrong' is a new thought in the unfolding discourse, but it contains anaphoric *this* as a 'retrospective signal', in Erdmann's terms, that would still fit in with the backward-looking character of intraposed subject clauses more generally. Interestingly, for the remaining *that*-clauses which do not contain such an anchor and are thus arguably "brand-new", the communicative effect is still similar: What these instances in the corpus have in common is that they achieve a kind of 'presupposition effect' in Kiparsky and Kiparsky's (1970) sense, "presenting the content of the non-extraposed clause as in

---

[31] For English, it was shown above that intraposed Gerunds are typically shorter than their associated matrix clause, but it is also true that they are shorter on average than intraposed *to*-Infinitives and *that*-clauses (6.4 words versus 10.7 words versus 11.4 words, respectively, cf. the means of the data provided by Kaltenböck (2004: 264)).

some sense 'given'"; it is "mainly a rhetorical device […] to persuade the listener to agree with the speaker's judgement by presenting it as a given fact." (Kaltenböck 2004: 277)

In sum, where there is synchronic choice, a strong factor motivating intraposition, and thus to override the effects of diachrony, entrenchment, weight and nominality, is discourse-pragmatics. As Sierwierska (1993: 840) puts it, "the fact that discourse considerations tend to outweigh purely semantic and syntactic ones as determinants of order is universally recognized." However, what we must not forget at the very end of this section is that there are still some SV/VO languages left that were presented as exceptions earlier on because they regularly (rather than exceptionally) place their finite subject clauses in situ, without any productive pattern of extraposition and without the analogical support of object clauses (i.e. the latter are postverbal). This concerned, for example, Chinese, Kana (A-clauses), Purépecha, Taba and Tetun. In view of such languages, but also on a more general level, one may thus wonder if there are even more basic requirements that need to be met before discourse-pragmatic motivations can assert themselves. I am asking this question because it has, in fact, been proposed that languages need to provide certain favourable conditions in order for the intraposition of finite subject clauses to be possible to begin with.

One such proposal of quite an interesting nature was made by Hawkins (1994): In the context of his processing theory mentioned repeatedly above, Hawkins' prediction is that weight effects basically motivate sentence-finality of subject clauses (i.e. as the typologically unmarked pattern for all subject-initial languages across the board).[32] However, in trying to motivate cases of intraposition in such languages (which go against his processing account), he notes that finite in-situ clauses in English only became attested at a certain historical stage, namely after the language had begun to grammaticalize a syntactic category of subject in many parts of its grammar. In Hawkins (1986), he had shown that, during the history of English, the subject (or {SA}) relation came to play an increasingly important role for the organization of syntactic processes such as constituent order, control, raising, long-distance dependencies, coreference in coordination, etc. Therefore, "the pressure presumably grew to call in all potential subjects under this expanding generalization", and "if we count sentential subjects as part of the larger category of subjects", this would motivate why these, too, ultimately succumbed to the strong {SA} attractor category and became possible in the preverbal position characteristic of subjects. (Hawkins 1994: 196) "The general prediction made by this account, of course, is that sentential subjects in subject position will be found only in languages with a strongly grammaticalized subject relation", and "one could presumably quantify the number of subject-sensitive rules in languages in order to test this" (ibid.). At the beginning of my research into subject clauses, this is precisely what I tried to do; as a result, the database contains a collection of 'subject-sensitive' rules for each language, including word order, case marking,

---

[32] This is formulated quite sweepingly in his 1994 book but modified by his later accounts, where SV/OV languages are argued not to show right-oriented weight effects, as we saw earlier.

different types of raising, control, relativization, etc.[33] However, it soon became clear that this information is extremely difficult to collect reliably and to quantify precisely. Hence a rigorous test of his prediction is impossible at this point. In regard to the 'exceptional' languages above, it may, however, be noted that his hypothesis yields mixed results. Chinese has been argued forcefully, by LaPolla (1990, 1993), not to have a syntactic category of 'subject' that is specifically 'isolated' by behavioural operations like raising, relativization, coordination etc.; all of these are syntactically unconstrained and do not specifically apply to {SA}-arguments only. Taba is similar to some extent (though probably not that radical), as far as I can tell. Purépecha and Tetun, by contrast, fare demonstrably better, as subjects control both coding properties (e.g. indexation, and case marking in Purépecha) as well as a number of behavioural constructions (cf., e.g. van Klinken (1999: 170) for discussion in Tetun). However, a simpler reason why these languages allow relatively sentential constructions in subject position was already given above: At least Chinese, Purépecha and Taba show a certain amount of 'left-branching' structures in their grammar, either as grammaticalized patterns (e.g. preverbal *ba*-objects and prenominal relative clauses in Chinese), or because there is considerably more word-order freedom than in strongly SV/VO languages (e.g. Purépecha). I consider it reasonable that the existence of such local analogical models for preposing various constituents exerts a more direct influence than the arguably more abstract analogy that subjects are targeted by different behavioural constructions. On the other hand, the fact that, for example, Old English used to have pervasive OV models in its grammar, left *that*-clauses completely unimpressed: As far as we know, they always occurred sentence-finally as subject (cf. Fischer and van der Wurff 2006: 171). An example is given below:

(127) Old English (Fischer and van der Wurff 2006: 167)

> *Hit ne   gerist nanum   ricum   cynincge* [*þæt   ... *].
> it   not befits no.DAT powerful.DAT king.DAT   **COMP**
> 'It does not befit any powerful king to …'

From this perspective, Hawkins' hypothesis for the later emergence of in-situ *that*-subjects in English seems to make sense: The earliest attestations of such patterns date to the Late Middle English period, after substantial changes to the word order and a certain grammaticalization of the subject position had been well on their way.[34] Therefore, in light of empirical pros and cons for Hawkins's proposal, the issue of whether a strongly grammaticalized subject relation favours the intraposition of finite subject clauses is still open to debate.

It should be pointed out that the historical data from English can also be interpreted from another perspective, one that brings us to a final possible precondition for intraposition. Fischer (2007) argues that the emergence of intraposed *that*-subjects is an effect of the incipient development of a written standard in Late Middle English.

---

[33] As was described in §3.3, this information can be found upon scrolling down to the lower right end of each entry in the database.

[34] Note that some works on diachronic syntax, such as Traugott (1972: 102), date the productive emergence of in-situ *that*-subjects much later, i.e. in the 19th century. At that time, further behavioural constructions such as control and raising had been well-established.

Her specific claim is that "languages which have developed a written standard undergo […] a 'Verschriftlichung der Sprache'" (Fischer 2007: 38), such that a more widespread use of certain structures in a common written medium may ultimately also influence the spoken language and lead to the acceptance of intraposition more generally. While this is as conjectural as Hawkins' proposal, there is substantial evidence that in English, the occurrence of intraposed *that*-subjects is tightly bound to the written language. It is telling, for example, that Kaltenböck's (2004) corpus does not yield a single instance of the pattern in spoken language; all cases discussed above come from written sources. This is confirmed by the conversational data analysed in Couper-Kuhlen and Thompson (2006: 28), and these authors, too, maintain that the carefully planned and highly systematic use of certain constructions in the standard written language wrongly creates the impression of a more general, orderly syntactic template in the grammar of the language (cf. Hopper 2004 on this point more generally). Spoken discourse, which unfolds under all kinds of contingencies and real-time pressures of language processing, is thus much more dependent, for example, on an optimal distribution of weight in the sentence (from both a production and a comprehension perspective), and hence the postverbal position of the complement is strictly adhered to.[35] It remains to be worked out by future research if this line of argumentation can also be extended to other languages with a firm and standardized written tradition. All that could be done here was to recognize it as a factor that has been discussed as a potential facilitator of intraposition.

We have now come a long way of describing the positional patterns of subject clauses and their organization in various types of languages, and we have drawn on an intricate and diverse set of potential explanations. In the following, I will try to summarize the findings again in the form of descriptive generalizations that capture the major types of pattern in my data. Such cross-linguistic statements are necessarily coarse-grained and conceal a vast amount of language-specific variation, but the ones that I am going to offer are still much more detailed than the positional generalizations proposed by earlier accounts.

Overall, the positional preferences of subject clauses confirm the hypotheses put forward by Dryer (1980): We have seen ample evidence for the 'final-over-initial' hypothesis in relation to both S- and A-clauses, and some suggestive evidence for the 'final-over-internal' hypothesis in regard to A-clauses. What emerges more specifically is that the position of subject clauses strongly depends on (i) the basic constituent order of SV and OV in simple sentences, (ii) the structural properties of the subject-clause construction, (iii) the positional behaviour of the corresponding (i.e. structurally identical) object clause. These conspire in the following ways:

---

[35] Couper-Kuhlen and Thompson's (2006) paper goes well beyond this, arguing that 'extraposed subject clauses' should not actually be seen as a syntacticized pattern of English grammar, but as a discourse routine which is syntactically and prosodically detached from its putative 'CTP'; the latter is regarded as a highly formulaic framing element that often has several successive independent clauses in its scope but does not contract a syntactic relationship to the clause that it allegedly 'subcategorizes for'. I personally feel that they overstate their case somewhat in this paper, but the general recognition of grammatical structures as communicative 'problem-solving routines' that are moulded in social interaction, is an important point (cf. Günthner 2009 for a similar perspective on *it*-extraposition in German). As stated in §2.2.2, however, this cannot be done justice to by a broad typological study of the grammar of complementation.

(128) Languages with SV and OV as their basic constituent orders in simple sentences tend to have in-situ subject clauses if the corresponding object clause also remains preverbal. This tendency appears to be stronger for A-clauses than for S-clauses, and possible explanations for the latter phenomenon include different weight differentials for S- and A-clauses and their matrixes, lexically-specific iconic effects in A-clauses, and information-structural differences between S- and A-clauses.

Generalization (128) accounts for the data in the following way. In 34 of the 42 SV/OV languages with subject clauses (= 81%), subject clauses preferably precede the verb and thus occur on the same side as phrasal subjects; the eight remaining languages show a mixture of different positioning types or a complete type shift. The majority pattern for this language type is thus that of preverbal subject clauses. Where individual subject-clause constructions allow for a certain flexibilization as compared to phrasal subjects, the corresponding object clause is normally at least as flexible (i.e. it is flexible or postposed, e.g. certain complements in Choctaw and Basque). And where there is a categorical type shift of the subject clause (e.g. certain complements in Persian, Rama and Supyire), the corresponding object clause shows the same postverbal pattern. Note that the reverse relationship does not hold: a postposed object clause in an SV/OV language does not automatically lead to a postposed subject clause: In Evenki and Somali, for example, object clauses tend to follow the verb, but the corresponding subject clause is preferably left in situ; in Hup, the object clause is flexible and the subject clause in situ – the above generalization is thus still valid. As many typological generalizations, it is not exceptionless, though, since we do find SV/OV languages with the potential for extraposing subject clauses while the corresponding object clause remains in situ. This was shown above for Wappo and Slave. However, it should be pointed out that both exceptions concern specifically S-clauses, for which (128) is relaxed anyway. It is, therefore, predicted to be harder that A-clauses follow while P-clauses remain in situ. But even here one counterexample can be found: In Choctaw, *cha-/-na*-complements do exhibit precisely this pattern.

The second generalization can be summed up as follows:

(129) Languages with SV and VO as their basic constituent orders in simple sentences tolerate desententialized subject clauses in situ, but can do this for sentential constructions only if they have a certain amount of analogical models for preposing structures (e.g. OV patterns), a strongly grammaticalized subject relation in their linguistic system, or lack a clause-initial subordinator. More commonly, then, finite subject clauses in these languages remain in the same postverbal position as the corresponding object clause. Where intraposition is possible in addition to the preferred postverbal position, it is typically motivated by information structure and may have a bias towards the written medium in the respective language.

This pattern is applicable to the rigid SV/VO languages we investigated in detail (e.g. Arabic, Hmong Njua, Mayogo, Ndyuka, To'aba'ita etc.), but it also covers the sentence-initial sentential structures in Chinese, Purépecha and Taba (preposing in the grammar), in English, Kana and Tetun (strongly grammaticalized subject relations) and Vietnamese (no initial subordinator, which also true for Chinese and Taba again).

Finally, it also extends, in certain language-specific shapes, to languages with SV and mixed VO/OV patterns, such as German, Ma'di and Hungarian.

The third generalization is the following:

(130) Languages with VS as their basic constituent order in simple sentences do not normally show preverbal subject clauses, unless this position is also available for phrasal subjects (Mapudungun) or unless the corresponding object clause is also preverbal (Barasano). In VSO languages, A-clauses do not usually appear in the centre-embedded position.

An overarching trend across all word-order types, and for both S- and A-clauses, is that desententialized and especially recategorized complements have at least as much potential to appear in preverbal position as sentential constructions. This suggests that we are dealing with a more fundamental principle of morphosyntactic organization here. As was mentioned before, the full extent of this principle is investigated in a larger project and will not concern us here. Instead, what we will turn to in the final section of the present chapter is a phenomenon that has cropped up at various points so far but deserves some comments in its own right: the 'place-holding' elements that can occur with subject clauses when they are extraposed.

5.5 Placeholders of subject clauses in the matrix

In his seminal overview of complementation, Noonan (1985|2007) notes that the 'right extraposition' of complement clauses (in his wide understanding of 'extraposition') may sometimes, though "not in all languages", be accompanied by a special "proform" in the matrix clause (Noonan 2007: 93). What Noonan primarily seems to have in mind is proforms with a relatively low referential status (e.g. personal *it* rather than a more clearly referential demonstrative such as *this*).[36] In the literature, such forms are variously known as 'placeholders', 'expletives', 'dummy' pronouns or 'anticipatory' elements.[37] They are illustrated by the following example:

(131) Koyra Chiini (Nilo-Saharan, Songhay: Mali; Heath 1999: 295)

A	*tilasu*	[*ni*	*ma*	*koy*].
3SG.SBJ	be.obligatory	2SG.SBJ	SBJV	go

'It is obligatory that you go.'

The distribution of such placeholders is rather straightforward in my data. Unsurprisingly, the languages whose postverbal subject clauses regularly occur with an

---

[36] The referential status of the anticipatory pronouns of subject clauses is a hotly debated issue. While formal accounts tend towards the position that *it* in English is "a purely formal syntactic placeholder […] devoid of meaning" (Seppänen 2002: 456–58), functionally-minded linguists have often aimed at a more referential and semantically richer interpretation (e.g. Bolinger 1977, Langacker 2009: ch.5). Surveying the arguments accrued in the literature, Kaltenböck (2004: 47) proposes a scale of gradience in the referentiality of *it*, with anticipatory *it* sitting right in between clearly referential uses on the one hand and so-called 'prop-*it*' uses (e.g. weather verbs) on the other. Demonstrative pronouns, by contrast, are firmly on the referential side.

[37] The language-specific literature shows that distinctions between these terms may be necessary in order to capture the subtle behavioural differences of an anticipatory element in different contexts. In German, for instance, the anticipatory pronoun *es* in extraposed subject clauses can have a 'place-holding' and a so-called 'correlative' ("Korrelat") function depending on the specific context (cf. Paranhos Zitterbart 2002: §3.1 for discussion).

expletive pronoun in the matrix clause are those in which subject pronouns are also required in simple sentences, i.e. languages without 'pro-drop'. This concerns German, Hausa, Koyra Chiini, Mayogo, Ndyuka, Supyire, To'aba'ita, and possibly Fongbe[38]. In other words, the distribution of expletives follows the distribution of the categories of pronominal subject expression in Dryer's (2011a) *WALS* chapter. Interestingly, though, there can be exceptions to this general rule. Hmong Njua, for example, normally requires overt pronouns in simple sentences, but postverbal subject clauses are *not* anticipated by a pronominal subject in the matrix, as can be seen by going back to example (120) from above. As discussed by Harriehausen (1990: 238), this is because the third-person singular pronoun *nwg* 'he/she/it' must be fully referential, i.e. in its neuter meaning 'it', it can only be used in referential contexts like 'I saw/did it'; less clearly referential uses such as a grammatical cataphor of subject clauses appear to be ruled out. Conversely, there are languages without obligatory subject pronouns in simple sentences which can show anticipatory pronominal elements for subject clauses. One example comes from A-clauses in Kana. This language normally expresses pronominal subjects by verbal affixes but does not cross-reference an overt subject NP (i.e. if there is a lexical NP, there is no verbal indexation). As the earlier example (115c) from above shows, this pattern is also adhered to for in-situ subject clauses, but when the same construction is extraposed (as in (115b)), a placeholder can optionally show up.

As with other grammatical markers, the optionality of placeholders is a particularly interesting phenomenon, although it is rarely made explicit in the materials what it is governed by. In Fongbe, for example, the predicate 'seem' takes a postverbal quotative clause as its complement and the expletive in the matrix is optional; other predicates (e.g. evaluatives like 'good') take a postverbal complement clause derived from a purpose construction, and here the expletive does not seem to be optional but obligatory (cf. (121) above for an example). It remains to be explored whether this has to do with the CTP as such or with the specific type of complement. In German, the situation is clearer: The expletive *es* "ist im Vorfeld obligatorisch, wenn keine andere Vorfeldfüllung vorhanden ist." (Paranhos Zitterbart 2002: 68) This is shown in (132):

(132) German (Indo-European, Germanic: Germany, Austria, Switzerland)

    *(Es)   *ist*   *gut*   [*dass*   *du*   *wieder*   *gesund bist*].
    3SG.SBJ   be.3SG.SBJ.PRS good   COMP   2SG.SBJ   again   healthy be.2SG.SBJ.PRS
    'It is good that you are healthy again.'

However, "tritt ein anderes Satzglied im Vorfeld auf, kann *es* (abgesehen von den Konstruktionen [...] *es+können/mögen+sein*) wegfallen" (ibid.):

(133) German (Indo-European, Germanic: Germany, Austria, Switzerland)

    Daher   *ist*   *(es)*   *gut*   [*dass*   *du*   *wieder*   *gesund bist*].
    therefore be.3SG.SBJ.PRS 3SG.SBJ good   COMP   2SG.SBJ   again   healthy be.2SG.SBJ.PRS
    'Therefore, it is good that you are healthy again.'

---

[38] In Fongbe, the relevant elements are clitics rather than pronouns but still grammatical words and not affixes on the verb.

This constraint, then, is due to the V2-syntax of present-day German declarative clauses. Interestingly, the same sort of argument may be brought to bear on earlier stages of English. Warner (1982: 78), for instance, observes that anticipatory *it* in Middle English is already very regular in his corpus (the *Wyclifite Sermons*), but in contrast to Present-Day English, there are certain contexts in which it can be left out. This happens precisely "when some other element (adverb, NP, PP […]) precedes the finite verb" (ibid.). It seems to me that this could be accounted for as a persistence effect of the V2-syntax of Old English. What happened in Middle English is that the general reorganization of the language towards the strong SVO template made it possible for anticipatory *it* to slowly encroach upon these V2-patterns, yielding variation in use in extraposed subject clauses:

(134) Middle English (Warner 1982: 78)

    a. *þerefore is nede hem to wite what dedis þat þei shulden do, ...*
       'Therefore, it is necessary for them to know what they should do.'

    b. *þerefore it were nede to hem to knowe witt of þes wordis, ...*
       'Therefore, it was necessary for them to know what these words mean.'

As can be seen, the V2-construction in (134a) coexisted with the expletive one in (134b), which later came to be conventionalized as the only pattern.

    Some further interesting observations can be made when the class of "proforms" that anticipate postverbal subject clauses is widened to encompass some more clearly referential forms, such as (grammaticalized) demonstratives. In Hungarian, for example, the distal demonstrative pronoun *az* is optionally inserted into the main clause in order to "signal the function of a [finite] subordinate clause" (Rebrus and Babarczy (n.d.: 379); note that Kenesei et al. (1998: 28) describe this morpheme as "an expletive pronominal *az* 'it'". The following example shows the occurrence of *az* in anticipation of an object clause, where one can see that it is case-marked appropriately:

(135) Hungarian (Uralic, Finno-Ugric: Hungary; Kenesei et al. 1998: 29)

    Anna   tudta   (azt),   [hogy Péter beteg].
    Anna   knew.DEF it/that-ACC   COMP Peter sick
    'Anna knew that Peter was sick.'

In subject complementation, too, it can anticipate a right-extraposed complement, but, interestingly, it can also yield structures such as the following:

(136) Hungarian (Uralic, Finno-Ugric: Hungary; Kenesei et al. 1998: 29)

    (Az-ø)   [hogy Péter beteg volt], mindenki-t   meglepett.
    it/that-NOM COMP Peter sick was   everyone-ACC surprised
    'That Peter was sick surprised everyone.'

What we are observing here is the immediate succession of the anticipatory element and the subject clause, without the latter being extraposed beyond the matrix clause. We would expect that the in-situ position of the subject clause renders the occurrence of an expletive superfluous. For English object clauses of the same type (e.g. *I regret it* [*that I didn't go to the exhibition*]), Bergh (1997) thus speaks of "vacuous extraposition". Bergh argues that this construction underlines the factual (i.e.

presupposed) character of the complement. Incidentally, exactly the same kind of discourse-pragmatic characterization has also been given for vacuous extraposition of finite subject clauses in Modern Hebrew (a language beyond the scope of the present sample):

(137) Modern Hebrew (Afro-Asiatic, Semitic: Israel; Glinert 1989: 335)

Ze [*she-zo ta'ut*]   *barur li.*
it   that-it mistake plain to.me
'That it's is mistake is plain to me.'

What we can see here is a parallel structure to (136) from Hungarian above: The "empty *ze* 'it'" (ibid.) occurs immediately before an intraposed subject clause. Glinert (ibid.) argues that the "inseparable" combination *ze she* at the sentence-initial position "signif[ies] 'the fact that'" and thus works essentially like an appositive construction. Although I have no explicit information on the discourse-pragmatic effect of overt *az* in the Hungarian structure in (136) above, it is nevertheless true that it does have the *structure* of an appositive construction (bearing in mind that we are actually dealing with a demonstrative pronoun here). And this is, I believe, a straightforward persistence effect of its diachronic origin. Recall from our earlier discussion that *hogy*-complements derived historically from relative clauses (just like German *dass*, Georgian *rom*, etc.), and the most likely bridging context was that of an appositive construction (cf. §7.2 for details). This is also true of Hebrew: Givón (1991) makes a detailed case for how relative *(a)she(r)* came to be a complementizer, and the essential context here was again that of a head noun like 'fact' which then came to be omitted, yielding a complement clause. It is telling that in German, Georgian and Hungarian, an expletive or demonstrative still regularly surfaces in the matrixes of finite complement clauses, and the sentence-initial succession of the pronominal element and the complementizer is just a special case of this more general pattern. And there is another interesting thing about it: The occurrence of *az* in Hungarian appears to correlate statistically with the position of the complement clause: "Geht der Nebensatz seinem übergeordneten Satz voraus, so werden in der Regel die palatalen Formen des Demonstrativpronomens [*az/azt*] als Deutewörter benutzt" (Szent-Iványi 1995: 137), as in (136) above and in the following examples:

(138) Hungarian (Uralic, Finno-Ugric: Hungary; Kenesei et al. 1998: 29)

a. *Az-t,   [hogy Péter beteg volt], Anna tudta.*
    it-ACC COMP Peter sick   was   Anna knew
    'That Peter had been sick was known to Anna.'

b. *Anna az-t,   [hogy   Péter beteg volt], tudta.*
    Anna it-ACC   COMP Peter sick   was   knew
    'Anna knew that Peter had been sick.'

The apparent relative frequency of these structures can be interpreted in such a way that the insertion of the demonstrative before the *hogy*-clause facilitates on-line processing because it signals explicitly how the subordinate clause needs to be integrated with the main clause and effectively creates an "endocentric" construction (Grosu and Thompson 1977: 144), i.e. one with an overt nominal head. This echoes

precisely what we saw earlier on: In the presence of an explicit nominal element as the 'head' or 'flag' of a complement clause, preverbal position is tolerated even where the complement is relatively long and thus violates potential weight considerations. In this respect, Hungarian is the 'performance equivalent' of Persian, for which we saw that its *ke*-complements *must* be anticipated by a nominal marker (lexical noun, demonstrative, article) if they are to occur preverbally (cf. (104) again); in Hungarian, it is possible for *az* to be omitted. Therefore, Hungarian, too, joins in – on a more subtle plane – the cross-linguistic preference for making preverbal complement clauses more overtly nominal than postverbal ones.[39]

An important thing to realize in the context of proforms is that they are, of course, not limited to anticipating an ensuing subject clause. The mirror image, i.e. anaphoric reference, is also attested. In SV/VO languages, this commonly happens when a normally postverbal finite subject clause is topicalized (depending on the precise morphosyntactic and discourse-pragmatic constellation, such processes may be called 'left-dislocation'). This was briefly mentioned for Hausa earlier on, and the same facts apply, for instance, to Serbo-Croatian. I stated earlier that my informant basically rejected finite in-situ subject clauses, and while linguistic experts on the language say that it may be a grammatical option, they also agree that it is still extremely rare (Wayles Browne, p.c.). Much more common is for sentence-initial subject clauses to be resumed by an anaphoric demonstrative in the main clause ('That Perišić played in the first team, *this* surprised me'). In Lao, one complementation pattern can function as an in-situ S-clause, but there is also a variant in which the subject position is filled by a third-person pronoun, with the complement topicalized into the "extraclausal Left Position" (Enfield 2007: 467). Compare the following two examples:

(139) Lao (Tai-Kadai, Kam-Tai: Laos, Thailand; Enfield 2007: 467)

    a. [*Vaw3 phaa2saa3 qang3kit2*], *man2* *kaø* *ñaak4.*
        speak language English **3BARE** TOP.LINK difficult
        'Speaking English, it's difficult.'

    b. [*Vaw3 phaa2saa3 qang3kit2*] *kaø* *ñaak4.*
        speak language English TOP.LINK difficult
        'Speaking English is difficult.'

According to Enfield (ibid.), the second pattern "provides a way to express the same idea with tighter grammatical cohesion", precisely by making the bracketed unit the actual subject of the predicate 'difficult' (cf. also Enfield's (ibid.) insightful discussion of the respective tree diagrams). 'Left-extraposed' structures such as (139a) are also found in SV/OV languages. An example comes from Tümpisa Shoshone:

---

[39] The fact that, in both Hungarian and Modern Hebrew, the correlative element (*az, ze*) is restricted to finite complement clauses and does not co-occur with the respective Infinitival complements, can also be interpreted in the same light: On the one hand, it may be a diachronic persistence effect: Only *hogy-* and *she*-clauses have an intimate association with a demonstrative, while the Infinitive arose in a completely different diachronic context. The synchronic distribution of demonstrative correlates may simply reflect this original association. On the other hand, support by a demonstrative is also not needed in many Infinitival constructions since they are already desententialized and thus better adapted to the subject position in these languages to begin with.

(140) Tümpisa Shoshone (Uto-Aztecan, Numic: USA; Dayley 1989: 375)

[*Ün*   *nanümi*   *na-maapüatu-nna*]   (*sutü*)   *tsawinnuh.*
your   relative.OBJ   PASS-help-INF:DS   (**that.NOM**)   be.good
'For your relatives to be helped is good.'

Infinitival subject clauses in Tümpisa Shoshone "most often precede the main clause verb, but they may also follow it" (Dayley 1989: 376). In either case (!), a Nominative-marked "demonstrative may optionally appear in the main clause as a pronoun copy coreferential with the subject complement clause" (ibid). This mirrors the structure of simple sentences, in which overt subject pronouns are also optional rather than obligatory. Nonetheless, the language has a strong 'resumptive' character overall: It is common, even in simple sentences, for "direct arguments, such as subjects and objects, [to be] repeated in the same clause", usually by a "pronominal demonstrative 'cross-referencing' a third-person argument" (ibid.: 21). The same structure is carried over to relative clauses, which often contain an anaphoric copy of the head noun inside the relative clause (but one that still marks the external rather than the internal role of the head NP, cf. ibid.: 359). It is thus not surprising that complement clauses, too, can have a double representation by an anaphor or cataphor in the matrix clause. Of course, the use of anaphoric pronouns as in (140) above may also be obligatory rather than optional, and this results in the 'left-adjoined' subject clauses that we saw at the beginning of this chapter (cf. (101) from Sanuma again). Therefore, what we are dealing with here more generally is a synchronic cline from left-adjunction to embedding in subject position.[40]

This, then, leads us to consider the syntactic status of the complementation patterns in question. As was argued in Chapter 2, genuine complement clauses are, by definition, embedded as an argument of a CTP. As an adjoined clause, they cannot play this role and must hence be considered complementation strategies. This point has also been made for right-extraposed complements. Noonan (2007: 94), for example, states that "extraposition normally has the effect of not only removing the complement from its grammatical position, but also of depriving it of its grammatical role": It is the place-holding element that functions as the grammatical subject or object of the matrix clause, so that the complement itself is adjoined to a syntactically saturated matrix clause. If taken seriously, this observation has important implications for the typology of complementation systems. Specifically, it makes all complementation patterns that are *obligatorily* 'extraposed' from the canonical in-situ position complementation strategies rather than complement clauses in Dixon's (2006a) sense. (Where in-situ positions are possible, there is evidence that the construction in question can function directly as an argument of a CTP, and hence it can be called a complement clause in at

---

[40] It is possible that, in some cases, this also corresponds to a diachronic cline from adjunction to embedding. In the domain of relative clauses, it has analogously been argued that right-adjoined subordinate clauses can become "attracted" to an NP in the main clause and reanalysed as an embedded modifier of that NP (Hale 1975), and that left-adjoined correlative clauses can become embedded as internally-headed relative clauses by virtue of the anaphor in the matrix clause being dropped (Lehmann 1984).

least those instances.[41]) In our assessment of complementation systems later on (cf. §6.4 and §7.3), we will thus have to take such positional questions into account again and reconsider carefully the range of constructions that can truly be said to be embedded as subjects of a CTP. The whole question is undoubtedly most relevant to complementation in Dixon's (2006a) framework, given that it was he who proposed a rigorous distinction between complement clauses (henceforth CC) and complementation strategies (CS). Interestingly enough, though, Dixon himself does not appear to be consistent here. Recall from Chapter 2 that he rigorously assigns CS-status to constructions that contain "an NP or pronoun or demonstrative" in the matrix clause (Dixon 2006a: 38); the alleged complement is said to be "appositive", with the antecedent in the matrix clause "having reference to the entire apposed clause" (ibid.). So far, so good. But in his discussion of the typological parameters of complement clauses, by contrast, he lists position as one variable to be investigated, and asks field workers to pose the following question: "Does the complement clause occur at the same position in the main clause as would an NP filling the same argument slot? May or must a complement clause be extraposed to the end of the main clause?" This formulation is puzzling because he still speaks of complement clauses in the context of extraposition, so this process does not appear to result in the loss of the CC status; and importantly, his wording "may *or must*" shows that it also comprises cases in which there is no intraposed alternative. Apparently, then, 'extraposition', in both its performance and its grammatical sense, is treated differently from what he calls 'apposition'. The difference may lie in the specific nature of the antecedent, i.e. whether it is truly expletive or more clearly referential, such as a lexical NP or a demonstrative pronoun. But as we saw above, this distinction is gradient at best (if not somewhat artificial in the context of complementation), and the fact remains that Dixon's CC/CS distinction is primarily based on syntactic criteria (i.e. on whether the complement is directly embedded as an argument or not). From the latter perspective, the semantic nature of the antecedent is irrelevant – in both 'apposition' (in Dixon's sense) and 'extraposition' (in Noonan's sense), the subordinate clause adjoins to a syntactically complete matrix and does not function as an argument, as we saw above. In sum, a consistent application of Dixon's proposed CC/CS distinction forces us to relegate obligatorily extraposed complements with placeholders to the arena of complementation strategies.

But the real complication is yet to come: As Dryer's (2011a) data demonstrate, by far most of the world's languages are such that they do not employ overt subject pronouns to begin with. In 437 of his 711 languages (61.5%), the primary means of expression of pronominal subjects is that of person indexes on the verb, so that free subject pronouns do not normally occur in addition ('pro-drop'). Moreover, of the 143 languages that do employ free subject pronouns rather than (or in addition to) indexes,

---

[41] This position is taken by Mair (1990) for English subject clauses. He claims that extraposed complements can be regarded as 'subjects' because they can normally be 'moved back' to the subject position, and because "from a logical and semantic point of view they are in the same relation to their superordinate predicate as their non-extraposed variants" (Mair 1990: 21). However, on a consistent application of this argument, one would have to say that some predicates in English, notably *seem*, *appear* and related ones, NEVER take clauses as their subjects because they require obligatory extraposition. This is precisely the stance adopted by Mair (ibid.).

61 (= 42.7%) allow for these pronouns to be optional. In other words, the great majority of languages leaves the subject position usually unfilled if its referent is a pronominal entity. This is significant for complementation because in those languages, a complement that does not have a coreferential subject pronoun in the matrix clause looks perfectly embedded, but it may also be argued to be coreferential with a pro-dropped anaphor or cataphor in the matrix clause. This problem was briefly mentioned in connection with Chinese above (please consult (117) again). Many grammatical accounts of complementation in Chinese (e.g. Yue 2003, Lehmann 2002) assume that an unmarked clause functions directly as the subject or object of a CTP. This is also stated in Li and Thompson's (1981) description, but as we saw above, they also do not rule out an interpretation in which the matrix clause is "without a subject", with "the understood subject refer[ring] to [a] proposition" (ibid.: 480). In other words, due to the morphologically unmarked nature of the complement, there is surface ambiguity between a directly embedded subject clause and one that is adjoined to a null-instantiated third-person anaphor. The problem does arguably not arise where a sentence-initial complement receives grammatical marking typical of subjects, such as a case marker. But where it goes unmarked (as in Chinese), or where the marking is not typical of subjects, we are in a certain 'grey area' between adjunction and embedding. This is illustrated best for incipient complementation patterns that derive from adverbial clauses. A nice case in point comes from Epena Pedee. In this language, the evaluative predicative adjective *pía* 'good' can occur without an overt subject if this is pronominal; compare (141a) and (141b) below:

(141) Epena Pedee (Chocó: Colombia; Harms 1994: 24, 176)

    a. *Hã kʰoráa pía bɨ.*
       that basket good be
       'That basket is good.'

    b. *Pía bɨ.*
       good be
       'That is good.'

If a whole proposition is to be evaluated this way, the 'complement' takes the form of an adverbial clause of cause:

(142) Epena Pedee (Chocó: Colombia; Harms 1994: 206)

    [*Éperã pá-da pérã*] *pía ba-hi-ʔe-kʰá?*
     person AUX-PST **CAU** good be-PST-NEG-Q
    'Was it not good that he had become a person?'

Just as in Chinese, it is not entirely clear whether the sentence-initial clause occupies the subject position directly (as a counterpart of (141a) above), or whether the matrix clause contains a pro-dropped subject pronoun (as in (141b) above) to which an adverbial clause is adjoined. Semantic considerations would lead one to assume the complement interpretation because, as the translation suggests, it is the content of the proposition itself that is evaluated as good; it is not the cause for something else that is good ('Because he had become a person, it/that is good'). In fact, there is some interesting diachronic evidence from other languages that causal clauses of this kind

have been reinterpreted as genuine complement clauses. Deutscher (2000: 41ff.) presents the case of Akkadian, where causal clauses introduced by *kīma* were reanalysed as complements. As Deutscher shows, the bridging context was low-transitivity verbs of speaking that allowed for both interpretations (e.g. 'I complained to the governor *kīma* (= because/that) the barley was not collected', cf. Deutscher (2000: 42)). When this pattern was analogically extended to genuinely transitive verbs ('see', 'hear'), an embedded complement clause emerged. Interestingly, the claim that it is now akin to an object argument can be supported by the fact that it acquired the ability to occur in passivization; as a result, the *kīma*-clause now also became available as a subject clause:

(143) Akkadian ([†]Afro-Asiatic, Semitic: Mesopotamia; Deutscher 2000: 54)

[*Kīma*	*tupšikk-um*	*šū*	*lā*	*labirta-šunu*]	*innamer.*
COMP	corvee.work-NOM	DEM:M.SG:NOM	not	old.debt-3M.PL.POSS	3M.SG.saw.PASS

'It was seen that his corvee work was not their old debt.'

Similar evidence for embedded complements that are ultimately derived from causal adverbial clauses is found in Jamiltepec Mixtec, for which Johnson (1988: 36) reports that one of most commonly used complementizers is "the conjunction *váytí* […] 'because'", which is also found in subject clauses:

(144) Jamiltepec Mixtec (Oto-Manguean, Mixtecan: Mexico; Johnson 1988: 36)

Ndīchā	[*váytí*	*chāha*	*ñā*	*shuhun*].
true	COMP	COMPL.give	she	money

'It's true that she gave money.'

In light of such attested shifts from adjoined causal clauses to embedded complement clauses, the assumption of an embedded subject clause in Epena Pedee is at least not implausible. But again, the principal difficulty remains: In pro-drop languages, the syntactic status of a complement at the sentence margin is hard to determine precisely.

This does, of course, also carry over to the right sentence margin. The rigidly postposed complements that we discussed at length above never appear in the subject position. In languages that regularly drop subject pronouns, it is unclear whether the postverbal complement clause stands in apposition to this implicit subject pronoun, or whether it is to be considered 'embedded' in the sense of acting itself as the subject argument of the CTP in question – one that happens to surface in a different position than a phrasal subject, for independent (e.g. weight) reasons. Examples of such constructions include our earlier (120) from Hmong Njua, or the following pattern from Lango:

(145) Lango (Nilo-Saharan, Nilotic: Uganda; Noonan 1992: 192)

Bèr	[*nî*	*rwòt*	*òmìò*	*lócà*	*dyàŋ*].
3SG.SBJ-good-HAB	COMP	king	3SG.SBJ-give-PRFV	man	cow

'It's good that the king gave the man the cow.'

As can be seen, pronominal arguments are indexed on the verb in Lango, which is why pronouns are not normally used in addition. But this leads precisely to the situation described above. A nice parallel case, which is mentioned here because of its explicit

discussion of the issue, is found in Goemai, a West Chadic language not included in the present sample. In this language, complement clauses introduced by *goepe* can be embedded as direct objects of a number of CTPs. In addition, they can also appear in environments of what we called 'experiencer-object' verbs in Chapter 2, such as the following one:

(146) Goemai (Afro-Asiatic, Chadic, West Chadic: Central Nigeria; Hellwig 2006: 215)

> *Sh'ang* *hen* [*goepe* *goe* *wul* *lu=noe*].
> pleasant 1SG.OBJ COMP 2SG.M arrive settlement=1SG.POSS
> 'It pleases me that you arrived at my home.'

Hellwig (2006: 214) comments on this example in the following way:

> It looks as if these clauses occur in S/A function, replacing a stimulus NP. However, I prefer to analyse them not as complement clauses, but rather as clauses in apposition, whereby the 3SG subject pronoun *ni* 'it' is omitted because it is recoverable from the context. [...] This analysis rests on the observation that the only criterion that reliably defines S/A is constituent order – but in ([146]), the clause in question does not (and cannot) occur preceding the verb.

This characterization echoes what I have been arguing: the possibility of pro-drop complicates the identification of the syntactic function of alleged 'subject clauses'. In analogy to the cases of obligatory 'right-extraposition' with an overt placeholder, it seems best to treat examples like (146) as being similarly relegated from subject position and function (cf. Noonan's quote again from above).

In my sample, a total of 27 of 138 S-clauses (= 19.6%), and 7 of 54 A-clauses (= 13%) is coded as involving such classificatory problems. In both S- and A-clauses, half of the relevant patterns are such that they contain an overt placeholder, while in the other half, the possibility of pro-drop leaves at least doubts as to the status of the complement in question. When we ask ourselves later on, 'What kinds of syntactic function can complements in language X occur in?', the resulting situation is thus that, for some languages, we will have to say, 'The S function can be covered semantically (such that a complement co-occurs with one-place CTPs) but not syntactically (because it cannot occupy the positional slot of phrasal arguments of the same CTPs).

In sum, the positional patterns of complements can have important implications for the syntactic status of the construction in question and thus impinge on our interpretation of complementation systems at large. With this in mind, we can now close the discussion of syntagmatic relations of complement clauses to their matrixes, and turn to the semantic or functional relations they contract.

6

Selectional relations to the matrix

6.1 Introduction

In this and the following chapter, I examine the co-occurrence patterns of complement clauses and different types of matrix predicates. In syntactic theory, these patterns have been described as 'selectional restrictions' that a matrix predicate imposes on the number of types of arguments it 'subcategorizes' for. For the purpose of cross-linguistic comparison, complement-taking predicates are usually arranged into semantically and/or syntactically coherent groups, which were established as 'CTP classes' or 'environments' of complementation in §2.5.2. The present chapter is devoted to investigating several aspects of the cross-linguistic distribution of these environments. The goal is again to provide a quantitatively orientated perspective on previous research into this issue, and to augment the latter with hitherto unstudied aspects of the topic. The chapter is organized as follows: In §6.2, I outline how the co-occurrence patterns of complements and CTP classes were determined and coded for the present study. In §6.3, I then go on to examine how productively the different CTP classes are covered by complementation constructions across the sample, and how this relates to their preferred morphosyntactic properties. In the typological literature, the latter issue was first captured systematically by Givón's (1980) 'binding hierarchy of complementation', and I intend to enrich the discussion of these form-function mappings by also taking a syntactic dimension into account, viz. the construal of the CTP as an S-, A- or P-taking predicate. We will see that the different syntactic environments, as well as their specific semantic subclasses, are not equally hospitable to complementation constructions. The bulk of the chapter is concerned with the documentation and motivation of these asymmetries. In the final section §6.4, the comparative syntactic concepts of S, A and P are then further dissected into canonical and non-canonical category members (e.g. direct versus oblique P-arguments), so as to examine the more specific syntactic environments in which complementation patterns come to function across the world's languages.

6.2 Determining co-occurrence patterns

The present section is the methodological preliminary to the analyses to come in this chapter and in Chapter 7. Conceptually, the central issue is how each of the complementation patterns in the database does or does not establish co-occurrence relations with the classes of matrix predicates set up in Chapter 2. For convenience, these are repeated here as an uncommented list (Table 17):

Table 17. Environments of complementation distinguished in the present study

Subject clauses	Object clauses
seem/appear-S	phasal-P
epistemic-S	causative-P
deontic-S	jussive-P
evaluative-S	same-subject desiderative-P
phasal-S	different-subject desiderative-P
other-S	knowledge-P
A-clauses	perception-P
	propositional-attitude-P
	emotive-P
	utterance-P

Retrieving the relevant information on co-occurrence patterns (and on the argument-structural properties of individual predicates in each language, to be examined in §6.4) was probably the most time-consuming aspect of the entire data-collection process. In regard to reference grammars, it required a good deal of the Greenbergian credo, "You gotta muck around in grammars" (cited in Croft 2003: 30): Rather than concentrating on the relevant chapters on complementation or clause combinations only, I looked for all kinds of relevant examples, structures and possible alternatives throughout the grammars. Where the materials were available to me in electronic, searchable, form, I performed exhaustive search queries for all potentially relevant predicates, translational equivalents, etc. The same holds for text collections accompanying the grammars. In addition, I benefitted from consultations with experts on and/or native speakers of the relevant languages, using tailor-made test sentences of the relevant complement-taking environment to elicit the translational equivalents in the language in question. These efforts notwithstanding, it has to be emphasized that the results must be seen as tentative and, where based only on search queries in reference materials, as no more than a first approximation of the actual situation. As a result, there are bound to be misclassifications of individual data points, but, as stated in Chapter 3, I am taking this risk here in order to make a first move towards measuring cross-linguistic distributional patterns in complementation. The relevant data are found on the lower left of the entry for each complementation pattern in the database (cf. the heading 'CTP distribution and functional potential', and the verbal comments on 'productivity' at the bottom). As can be seen, each of the environments has been given a categorical label, to be explicated shortly, and the 'comments' box below these categories typically justifies the chosen values in some detail. The purpose of the present section is now to explain the coding procedure that underlies the values to be seen in the database.

When dealing with selectional restrictions in complementation, the essential distinction is, of course, whether a given complementation pattern can or cannot occur with a CTP class in question. In German, for example, the Infinitival complement has to be chosen for same-subject desiderative contexts (*Ich will gehen* ('I want to go')), while the sentential *dass*-complement covers different-subject desiderative contexts (*Ich will, dass du gehst* ('I want you to go')). Swapping the two constructions and environments results in ungrammaticality, so they are in complementary distribution here. Such situations are very frequent in complementation systems, i.e. we know that a given pattern co-occurs with only a subset of the relevant CTPs of a language. However, a simply binary distinction does not do justice to the data, for several reasons. First, there are often gradations of the co-occurrence patterns, i.e. different degrees of likelihood by which two competing constructions are chosen by a given CTP. In Lango, for instance, it is possible (i.e. grammatically licensed) for same-subject desiderative predicates to choose a finite complement with Subjunctive mood, but this is exceedingly rare as compared to the regular choice of the Infinitive. In order to incorporate this kind of information (where available), some subtle gradation should be made at the 'upper end' of the co-occurrence possibilities. Second, it is sometimes observed that the co-occurrence of a CTP and a given complement is allowed only under certain semantic or grammatical restrictions (more on this below), so that here, too, a binary distinction into 'possible' and 'impossible' combinations is somewhat misleading. Finally, towards the 'lower end' of the co-occurrence scale, there is an intuitive difference between complements that are disallowed to occur with a given CTP class (because a competing construction in the system does that job) and those that never have a chance to occur in a certain environment because that environment is missing from the language altogether. For example, where a language employs desiderative affixes rather than a lexical verb 'want', there is no chance for a given complement to occur in such an environment. But in contrast to the first case, this is not an inherent distributional restriction on the complement as such (e.g. due to its structure or semantics), but a question of how the lexicon and the complementation system are organized.

Taking all of these considerations into account, we can distinguish between several co-occurrence types of complements and complement-taking environments. These are set out in Table 18.

Table 18. Co-occurrence patterns of complements and complement-taking environments

Co-occurrence type	Characterization	Grammaticality score
Full coverage	The complement is grammatically compatible with and productively used in the complement-taking environment.	1
Possible but unproductive	The complement is grammatically licensed in the complement-taking environment but strongly dispreferred to an alternative, more idiomatic complement.	0.9
Likely (but no attested instance)	The co-occurrence of the complement and the specific environment in question is very likely (e.g. because the relevant predicates are known to exist) but there are no attested instances of it in the materials available.	0.8
Partial coverage	The co-occurrence of the complement and the specific environment in question is subject to certain grammatical or semantic restrictions (cf.	0.5

	below for details).	
Not attested as a CTP	The environment in question is either not attested in the materials ('no information') or rendered by alternative expression formats outside the domain of complementation.	0.1
Impossible	The complement is ungrammatical in the environment in question, typically because a competing construction can complement the CTP in question.	0

The last three categories in Table 18 deserve some further comments. The distinction at the lower end of the spectrum can be made more transparent if we recall a graph used in Chapter 3, repeated here for convenience:
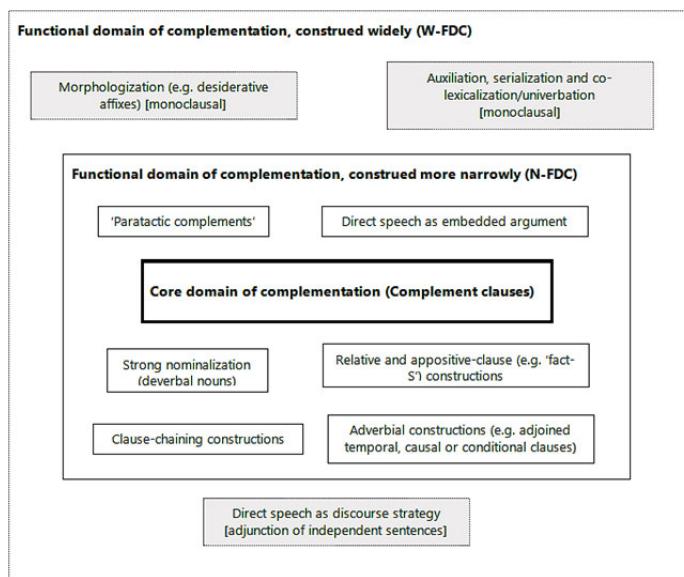
Figure 12. Expression formats for complementation

When we say that a certain environment is 'not attested as a CTP', this typically means that the relevant meaning is expressed by techniques outside the *narrow* functional domain of complementation as defined in Fig. 12, i.e. by morphological structures, verb complexes in simple sentences (auxiliation, serialization, univerbation), direct reported discourse that is not embedded as an argument of the predicate in question, or by altogether separate sentences ('He hit his wife. That is bad.'). In other words, the relevant expression formats are found at the fringes of Fig. 12, and they have in common (except for direct speech) that there is no lexical matrix verb of the given type that the complement could co-occur with to establish a biclausal structure.[1] What is also included in this category, however, are all cases for which no evidence could be

---

[1] The formulation "of the given type" is important here. Some languages do have, for example, desiderative lexical verbs, but they can only be used in reference to one's own actions. Different-subject desideratives ('I want you to leave') have to take a completely roundabout expression that does not involve biclausal constructions. In such cases, then, the CTP class as defined for the present study, i.e. the specific lexico-grammatical constellation, is not available to begin with, and hence no complementation structure would have a chance to occur with it. A similar situation arises when a certain semantic type of CTP (e.g. 'phasal predicates') is lexicalized as an object-complement structure (i.e. a 'phasal-P' class in our schema) and is never construed impersonally. This means that our 'phasal-S' class is not lexicalized in the language and hence 'not attested as a CTP'.

found for the CTP class to begin with. Despite intensive searches, this situation did arise in quite a few languages. It may be due simply to a gap in the available materials (i.e. the CTP does exist but is not documented in the grammar). In many cases, however, the description of the complementation system is fairly elaborate, detailing the CTP classes and syntactic functions that a given complement can fill, and if some of our specific environments are left completely unmentioned, it is indeed likely that they are not lexicalized as CTPs, but find alternative expression in the language. Of course, another possibility is that the authors of the reference materials did not encounter the relevant CTPs in their original corpora, which does not rule out that they exist. It is precisely because of this amount of uncertainty involved that these cases are kept separate from the ones where we know that certain CTP class exists but is ungrammatical with a given complementation pattern. This brings us to the category called 'impossible' in Table 18 above. I reserved this value largely for situations in which a given CTP class is lexicalized in the way intended (e.g. there is a lexical verb 'know' which takes the experiencer as subject and co-occurs with a clause spelling out the theme as non-subject), but it cannot select the complementation pattern in question because an alternative biclausal construction is earmarked for that. This competitor is typically another complement clause from the core domain, but it can also be one of the other biclausal constructions surrounding the core in Fig. 12 above (i.e. a complementation strategy). Note that the competitor may or may not itself be a data point in my sample; it was outlined in §3.2 that this depends on the amount of information available on the pattern in question and on whether it is a major or rather marginal complementation technique in the language. At any rate, the coding takes into account that there is a clear competitor for the CTP in question.

The middle category 'partial coverage' also requires some comment. The category is supposed to acknowledge the fact that a given complementation pattern can be well-distributed in general, but is applied to certain CTP classes only under specific circumstances. This, in turn, makes the classes in question different from the others because they exhibit limited rather than full coverage by the complement. Such restrictions can be due to a number of reasons, the most prominent ones of which are the following:

(i) In one case, a complementation pattern comes with a particular subordinator (obligatorily or optionally) but a restricted set of CTPs must leave the marker out, with the rest of the construction remaining structurally the same pattern. In §3.2, we discussed precisely this issue and mentioned the finite complement type in Noon, where the subordinator *an* must be omitted in same-subject desiderative environments. The German Infinitival complement is similar in that its subordinator *zu* must be left out in same-subject desiderative, causative and perception contexts (*I sah sie *zu gehen* 'I saw her *to leave'); in Amele, the same restriction affects perception and phasal-P contexts.

(ii) The opposite pattern is one in which a specific CTP class must add something to the complement or change the subordinator in order to take the constructional pattern at hand. In Basque, the highly productive nominalization construction needs to be enhanced by a locative marker before it can occur with phasal and perception verbs. (The resulting form *-t(z)en* has been described as "Imperfective Participle" (Trask 1997: 215), which Trask notes is the transparent combination of the nominalizer *-t(z)e* and

Locative -*n*.) In Wolaytta, one type of complement is flagged by a clausal nominalizer -*ga* and an appropriate case suffix; this is the Absolutive for most object functions, but phasal-P contexts require the Ablative instead (Lamberti and Sottile 1997: 239). Therefore, rather than saying that phasal verbs require a completely different kind of construction, it is acknowledged that they can appear in the same basic structural pattern but need a different marker. What is also included in this category are the cases (mentioned in Chapter 4) in which a nominalized complement has a more verbal ('infinitival') and a more nominal ('verbal noun') variant. The verbal one, being a proper complement clause, is taken as the basis for all coding decisions, but it needs to be acknowledged that some CTP classes may be compatible only with the verbal-noun variant. This holds, for example, for perception verbs in Noon. Consequently, they were coded as being restricted in their co-occurrence with the nominalized complement.

(iii) There may be semantic restrictions that a complement imposes on certain CTP classes. In Martuthunira, for instance, a 'finite future-oriented' complement can occur with propositional-attitude verbs, but only if they relate to a future event and include same subjects ('I think I'm going to die' but not *'I think she is going to die' or *'I think I've lost my wallet', cf. Dench 1995: 256f.). This is clearly a restriction of the propositional-attitude environment as opposed to other languages (and other complements in the same language). In Hmong Njua, the sentential complement marked by *kuam* is typically chosen in different-subject constellations and implies that the subject participant of the complement has a low degree of control over the realization of the dependent state-of-affairs (Harriehausen 1990: 219). Consequently, the construction can be extended to same-subject desiderative environments only if they also imply a low degree of control. Therefore, a sentence like 'I want to go' in the *kuam*-construction can be used only in contexts that imply 'it is not sure whether they will let me go' (ibid.). This too, was coded as partial coverage of the CTP class.

(iv) Partial coverage was also assigned to CTP classes whose co-occurrence with a certain type of complement met with mixed responses from native informants. My informants on Japanese, for example, showed a 50:50 split as to whether they accepted *koto*-complements in epistemic-S function ('It is likely/probable that …').[2] For Huallaga Quechua, Weber (1994: 101) reports that his informants found evaluative-S uses of certain complements extremely odd in elicitation (while they do occur – albeit very rarely – in his corpus).

(v) Finally, partial coverage was felt to be appropriate for constructions whose syntactic structure is not entirely clear. This will be illustrated later on in this chapter for so-called 'non-canonical' subject clauses, for which it is sometimes difficult to determine whether they represent a genuine syntactic subject or not.

There are further, language-specific, reasons for CTP restrictions, but the ones just outlined represent the major recurrent types. Metaphorically speaking, one can think of such restrictions as 'hiccoughs' in the lexical diffusion of a complement, and these will be significant when we perform dissimilarity analyses on the distributional data later on.

---

[2] I would like to thank Toshio Ohori and Kyoko Maezono for discussing the CTP distribution in Japanese with me, and especially again Kyoko Maezono for translating the relevant test sentences and sending them out to four further native speakers to elicit their judgements.

Returning to Table 18 again, one can see in the last column that the different co-occurrence types cannot only be coded in a categorical fashion. In addition, they could also be conceived of as a cline of 'grammaticality', i.e. of how acceptable a given combination of CTP class and complementation pattern is in a given language. It has variously been pointed out in the linguistic literature that grammaticality is not a binary but a gradient phenomenon (e.g. Schütze 1996, Hayes 2000), which is ultimately influenced by the frequency with which certain lexical and grammatical material has co-occurred in a language user's experience (cf. Bybee 2010 for several pieces of evidence). It is this stochastic information about co-occurrence patterns that is supposed to be reflected by the 'grammaticality score' in Table 18. However, in the absence of direct corpus frequencies, it can only be a very rough approximation of the data. It is included here at all because for some analyses, it will be useful to have a numerical indication of the **productivity** of a given complementation pattern and of a certain complement-taking environment. In relation to the former, productivity is understood here (again following Bybee's work) as the type frequency of a complementation pattern, i.e. the number of distinct environments it can be used in.[3] This can be expressed, and compared across languages, by adding up the individual grammaticality scores to a '**cumulative index of productivity**' (**CIP**) for the pattern in question. A highly productive complement, such as the *xa*=complement in Mixtec, is grammatical and idiomatic in many different CTP environments; as a result of this repeatedly high grammaticality score, it obtains a high CIP. This is illustrated schematically in Table 19 below:

Table 19. Co-occurrence matrix of complements and complement-taking environments

	Epis-S	Eval-S	Deon-S	…	SSwant-P	Know-P	Prop-P	Quot-P	CIP_{pat}
Mixtec-*xa*=	1	1	1	…	1	1	1	1	16.1
Mrt2	0.1	0.1	0.1	…	1	0	0.5	0.5	4.0
Toq-*na*	0.1	1	1	…	1	1	1	0.9	9.9
…	…	…	…	…	…	…	…	…	…
CIP_{class}	63.9	117.9	71.8	…	90.7	102.7	105.9	101.9	

The last row of the table, i.e. the column totals, indicates that the same kind of calculation can also be applied to the environments themselves: As we shall see shortly in §6.3, the CTP classes distinguished in the present study differ significantly as to how productively they co-occur with complementation constructions across the sample languages (or else choose entirely different expression formats, such as monoclausal alternatives). Here, too, then, a numerical indication of productivity will be helpful for cross-linguistic comparison.[4]

---

[3] Put differently, it measures the degree of lexical diffusion of a complementation pattern, but not on the level of individual lexemes (CTPs) but rather lexeme classes (CTP classes). Therefore, productivity as measured here thus makes no claim, for instance, about how many distinct evaluative-S predicates ('good', 'bad', 'difficult', etc.) there are, i.e. about the size of the class in a given language.

[4] I am aware of the fact that the numerical values in Table 18 are completely arbitrary. It is unlikely, for instance, that a difference of 0.1 on my scale reflects the real frequency differences between two competing constructions. What I intended to capture here for cross-linguistic comparison is that the difference between the two categories at the top of the scale is much smaller than their difference to the others; after all, both constructions are perfectly

Before we proceed to the analyses, a final preliminary needs to be dealt with. This relates to the fact that a given complementation pattern may be productive beyond the environments that happen to be selected for the present study. As was described in Chapter 2, the complement-taking environments studied here are frequently recurring constructional meanings across the world's languages, but they represent only a subset of the predicate classes that can be lexicalized for the purpose of complementation. This is especially true in the domain of object clauses: On the one hand, we saw in §2.5.2 that a number of constructional meanings in this syntactic domain had to be disregarded from the outset (due to limited cross-linguistic documentation or applicability). On the other hand, what can be observed very frequently in my data is that a certain matrix predicate is grammatically compatible with a given complementation pattern, but the resulting constructional meaning does not match my definition of the CTP class. Let me give two recurrent examples. Knowledge predicates were defined earlier as encoding declarative ('factual') knowledge when they combine with an appropriate complement. This may not rule out, however, that another complementation pattern in the same language can also combine with 'know', but to yield a different meaning, such as procedural knowledge (e.g. *I know how to cook vegan dishes*). For the coding procedure, this means that the latter construction is ungrammatical with the 'knowledge' environment studied here, but this conceals that it is, in fact, a somewhat more productive pattern than the coding suggests. Similarly, perception complements were defined here as encoding situations of immediate perception. When a complementation pattern is capable of combining with 'see' or 'hear', but only to render a propositional (hearsay or deductive evidential) meaning (e.g. *I've heard that the dean resigned; I saw* (= 'realized') *that someone had been drinking from my bottle*), this, too, makes it ungrammatical in the environment intended here but productive beyond it.

In order to capture such productivity effects outside of the predicate classes distinguished here, I added a final environment to the data, collectively referred to as '**other-P**'. In analogy to the 'other-S' class set up for subject clauses, this is basically a holistic waste-paper category that does not make many internal distinctions. Whenever there was evidence that a given complement is used productively beyond the criterial CTP classes, it was given the value '1' (i.e. 'covered') in the coding process; by contrast, if no such additional productivity could be established on the basis of the available materials, the value '0.1' ('not attested') was assigned. The category will not be examined qualitatively in the present chapter; it will become more relevant in Chapter

---

grammatical with a given CTP. And the same applies to the two categories at the bottom: Either scenario results in the absence of a co-occurrence and hence in a limit on the productivity of the complement. The difference here is meant to reflect that only the latter category imposes a real ungrammaticality constraint, while the last-but-one does not do so necessarily: In many cases, a certain type of co-occurrence is not attested in the materials, but it is not ruled out either. This category is, of course, problematic from a mathematical (or rather information-theoretic) point of view, but it will not be used for elaborate statistical model building, anyway. Some analyses will work straightforwardly with the categorical information (without transforming them into numbers), while the ones that do use the numerical information mostly involve non-parametric statistics, which treat the scale as ordinal rather than continuous data to begin with. This does not solve the information-theoretic problem, but it at least avoids implying meaningful distances between the categories. I would like to thank Christoph Rzymski for sharing his statistical expertise on these issues. He is not to be blamed for any errors in my analyses.

7, where it enters the productivity calculations for individual complementation patterns.

With these methodological considerations in mind, we can now proceed to the analyses. As was stated in the introduction above, the analyses of the present chapter will approach the co-occurrence data from the perspective of the CTP classes, aiming to characterize their cross-linguistically typical distributional properties. In other words, we will be examining the different *columns* (rather than rows) of Table 19 above.

6.3 Productivity and coding preferences of CTP classes

6.3.1 Overview

On a construction-specific level, the starting point is simply the observation that some environments of complementation are 'populated' over and over again by complementation patterns in my data, thus acting as powerful attractors for being coded by complement clauses, while others exhibit a much more selective behaviour, attracting significantly fewer constructions. On the level of languages, this results in the fact that, for a given language, some of the environments may not be rendered at all by complementation constructions, but are 'left blank', as it were, to be expressed by grammatical means outside of the (core) domain of complementation. A major goal of the present section is thus to establish these different patterns of productivity empirically, both with regard to individual CTP classes and the higher-level clusters of S, A and P. From a qualitative perspective, the results can be related quite naturally to Dixon's (1995, 2006a) observation that complement-taking predicates differ in how close their denotations are to prototypical 'verbal' (i.e. action/process) concepts. Dixon thus proposes to distinguish between 'primary concepts' and 'secondary concepts'. In Dixon (2005: 96), he writes that primary concepts are "those directly referring to some activity or state", so that the corresponding verbs "can make up a sentence by themselves with appropriate NPs." These are, for example, actions like motion, transfer, creation, mental activity or speaking. 'Secondary concepts', on the other hand, include "'not', 'can', 'try', 'want', and 'make'. [These] cannot be used by themselves but must be linked to a [primary concept], which is either explicitly stated or understood from the context". (Dixon 2006a: 11). In other words, they provide "semantic modification of some other [concept]." Admittedly, this is not a particularly satisfying definition, and there are undoubtedly problems in the details of this classification, but if we simply adopt it for the present moment, it will be interesting to see how much empirical support can be adduced for the distinction.

The conceptual content of complement-taking predicates is thus claimed to be related to the productivity of CTP classes: typically verbal concepts are hypothesized to be more likely attractors for complement clauses than 'secondary' concepts. The latter, Dixon proposes, will often be realized by verbal affixes, auxiliaries and forms of univerbation, i.e. monoclausal techniques. Even though Dixon does not make the connection explicit, the fact that morphosyntactic coding ultimately correlates with the conceptual content of complement-taking predicates is, of course, well-known: The

more tightly the matrix event and the complement event are conceptually integrated with one another, the more prone they are to diachronic developments which reduce the complementation construction to a monoclausal, syntactically integrated, entity. Dixon's secondary concepts would thus be expected to surrender to such developments, resulting in the fact that they are synchronically less productive as genuine complement-taking environments than primary concepts. This is, of course, nothing but the thrust of Givón's (1980) famous 'binding hierarchy of complementation'. According to this principle, different environments of complementation (i.e. CTP classes or specific uses of CTPs) show markedly different preferences of morphosyntactic expression, which can be arranged as a systematic scale of increasing or decreasing structural integration with the matrix clause.[5] Ransom (1986) and especially Cristofaro (2003) present empirically extended and theoretically revised versions of these form-function relationships, and the issue itself is too well-investigated to be taken up at length in the present study again. But to the extent that structural integration has a bearing on the productivity of a CTP class, it will be instructive to make reference to binding issues here. Moreover, given that – in contrast to all previous studies – I included a difference between subject- and object-taking environments in the definition of CTP classes, it will be interesting to see where the subject-taking environments are located on the binding scale.

In the following, I am going to integrate these different strands of research: the various environments of complementation will briefly be examined in relation to their cross-linguistic productivity, their alternative expression formats and their location on the binding hierarchy. The results are foreshadowed in Table 20 on the next page, which the entire discussion will draw on.

---

[5] The term 'binding' itself refers to the semantic principles underlying the form-function correlations, but since it has come to be so closely associated with the form-function interactions as such, I will use the term in reference to the latter here. In other words, my use of the term 'binding scale' relates exclusively to the hierarchical coding differences between various CTP classes and *not* to the semantic factors of 'binding' underlying them.

Table 20. Environments of complementation: Productivity and morphosyntactic coding

	Seem-S	Epis-S	Deon-S	Eval-S	Phas-S	Other-S	A	Phas-P	Caus-P	Juss-P	SSwant	DSwant	Perc-P	Know-P	PropAtt	Emot-P	Quot-P
Covered	17	44	49	100	12	66	47	58	40	95	76	73	91	94	90	84	88
Possible	1	0	2	2	0	1	0	1	12	13	7	6	12	5	8	2	10
Likely	0	4	7	13	3	10	10	0	1	3	0	2	4	1	1	12	0
Partial	5	7	9	6	0	5	7	6	3	8	7	8	8	3	11	6	7
Not att.	177	132	109	27	203	102	119	84	133	21	49	54	10	19	24	68	14
Imposs.	28	41	52	80	10	44	45	77	39	88	89	85	103	106	94	56	109
CIP$_{pat}$	38.1	63.9	71.8	117.9	34.7	87.6	70.4	70.3	66.4	115.1	90.7	89.4	110	102.7	105.9	105.2	101.9
CIP$_{lgs}$	33.9	58	65.9	121.2	23.5	85.1	65.4	69.1	59.7	115.8	89.5	87.7	112	102.5	105.8	102.6	102.9
Concept	sec	sec	sec	mixed	sec	mixed	prim	sec	sec	prim	sec	sec	prim	prim	prim	prim	prim
DepV	0.333	0.438	0.431	0.513	0.733	0.571	0.579	0.729	0.321	0.351	0.506	0.346	0.402	0.31	0.252	0.489	0.173

Binding scales of complementation

My data	Givón 1980	Cristofaro 2003
Utterance.P	Utterance	Utterance
		PropAtt
Prop.Att.P	PropAtt	Knowledge
Knowledge.P	Knowledge	
		Perception
DS.want.P	Desiderative	
Caus.P		Desiderative
Jussive.P	Jussive	Jussive
Perception.P		Causative
Seem.S		
Epis.S	Desiderative	
Eval.S, Deon.S, Emot.P	(same subject)	Phasal
SS.want.P	Causative	
	Phasal	
Other.S		
A		
Phasal.P		
Phasal.S		

	seemS	episS	deonS	evalS	phasS	otherS	A	phasP	cauP	jussP	SSwantP	DSwantP	percP	knowP	propattP	emotP	quotP
CIDmean	2.19	2.43	2.54	2.54	3.9	2.97	3.06	3.64	1.93	2.02	2.67	1.85	2.06	1.59	1.43	2.55	1.01

dependentV   independentV

Let me explain what the individual rows of the table are supposed to illustrate and how the results were obtained. The core of the table lists the individual environments of complementation (= 17 CTP classes) as columns. For each class, the first six rows provide an overview of the different co-occurrence frequencies with a complementation pattern, in absolute counts: For example, how often is the class in question fully covered by a complementation pattern in the data (row 1), how often does it occur with restrictions (row 4), and how often does it fall into the 'not attested' class as defined above (row 5)? In each environment, the figures of the first six rows thus always add up to the total number of complementation patterns in the database, i.e. N_{pat} = 228.

The next two rows then calculate, based on the matrix developed in Table 18 above, the cumulative index of productivity (CIP) for each CTP class. This is first done on the level of constructions (row 7), where the figures from the first six rows are multiplied with their corresponding grammaticality scores from above (Table 18). For example, the column on utterance predicates ("quot-P") would get a CIP of $((88 \cdot 1)+(10 \cdot 0.9)+(7 \cdot 0.5)+(14 \cdot 0.1) = 101.9)$. The resulting figure is thus an aggregate value that reflects how often a given CTP class establishes the different co-occurrence types with the complementation patterns in the data (e.g. 88 times fully covered (= $88 \cdot 1$), 7 times partially covered (= $7 \cdot 0.5$), and so on). On the level of languages, the CIP in row 8 records basically the same information, but from the organizational perspective of a linguistic system: The question here is whether a given language expresses the environment in question with a complementation construction (yes = 1, possible but unproductively = 0.9, likely = 0.8, partially = 0.5) or not (or not in the available materials) (= 0.1).[6] It is this index that can ultimately give us an idea about which environments are more prone to be rendered outside of the domain of complementation than others. And since we saw above that this may be related to Dixon's distinction between 'primary' and 'secondary' concepts, the respective concept type for each environment is also included in Table 20 (row 9); the decisions here are mostly taken directly from Dixon (2006a: 9–14; but cf. below for the subject classes).

The final rows provide two measures of the preferred morphosyntactic structure of the complements in each environment. In row 10, I count how often the complements that co-occur with a given CTP class include a dependent as opposed to an independent verb form. This distinction was treated in detail in §4.2, and what is presented in Table 20 is the proportion of dependent verb forms that characterize the

---

[6] As can be seen, the numerical values are largely the same as the ones for the CIP calculations at the construction-specific level, but they have now been aggregated to arrive at a CIP for each environment in a language as a whole. For example, if a language has two complementation patterns in my data and both of them are fully compatible with jussive CTPs, the jussive environment would get an overall CIP of 1+1 = 2 in this language. If quotative environments are fully covered by one the two complements but ungrammatical with the other, the overall CIP for utterance CTPs would be 1+0 = 1. If an environment is generally not attested (i.e. known or likely to be expressed by means outside of the domain of complementation), its CIP remains 0.1 for the entire language. And so on. The calculations were not trivial, though, because I had to take all biclausal constructions into account, not just the ones recorded in the database. It is possible, for instance, that a given CTP class is not covered by the complementation patterns that entered my data, but that it is still rendered by a complement or an alternative biclausal construction that may easily be reanalysed as a complement (relative clause, adverbial clause, clause chaining, etc., cf. Fig. 12 again). In the former case (i.e. coverage by a complement), the overall CIP is, of course, 1; in the latter case (i.e. coverage by an alternative biclausal construction), the CIP assigned is 0.5 ('restricted coverage').

complements of each class; this is first given numerically (row 10) and then followed by a graph that visualizes these proportions (row 11).[7] As can be seen at a glance, the different CTP classes do show markedly different proportions of dependent verb forms, so that the distribution as a whole is significantly skewed (randomized $\chi^2 = 98.65$, B = 100,000, $p < 0.0001$). And a similar skewing can also be observed if we choose a different, more comprehensive, measure of morphosyntactic downgrading, i.e. the cumulative index of desententialization (CID) developed in Chapter 4. Row 12 provides the mean CID for each environment, i.e. the mean degree of desententialization of the complements that co-occur with each class. In the final row of Table 20, we then find a boxplot display of these different CIDs. The graph visualizes the different environments in ascending order of their mean CIDs, from utterance predicates on the very left to phasal-S environments on the very right. Note that the respective mean cannot be read off the graph directly: The bold black lines in each box represent the *median* value of the CID in each class. The boxes, the whiskers and the outliers (i.e. the dots above the end of the whiskers) provide an indication of the dispersion of the CIDs in each group, i.e. of how widely the individual CIDs vary around the mean in each environment. The data in this graph, too, can be submitted to statistical analysis. An appropriate non-parametric technique for comparing the central tendencies of several different groups is the Kruskal-Wallis test (cf. Field et al. 2012: 674ff.). Being non-parametric, it considers the medians rather than the means of each environment, returning a highly significant signal for the distribution as a whole ($H = 111.8$, $df = 16$, $p < 2.2e\text{-}16$). The individual contrasts that contribute to this overall signal will be discussed in the sections to come.

What both graphs corroborate is that the different environments of complementation are associated, in a statistical sense, with different morphosyntactic properties of the complement. As was mentioned above, it has been customary in the typological literature to speak of 'hierarchies' of the morphosyntactic integration of the complement into the main clause: The more desententialized the complement, the less autonomous its status as a separate clause, and the closer the grammatical bond with the matrix clause. Accordingly, the different CTP classes can be arranged on a scale that reflects the relative degree of desententialization of their complements. As explained in footnote 5, I will refer to this scale as the 'binding hierarchy' of complementation, and on the right-hand side of Table 20, I have included three versions of this hierarchy: Givón's (1980) original proposal (though with adapted terminology[8]), Cristofaro's (2003) revision based on a thorough empirical basis, and, finally, a scale derived from my own data. The latter is conceptually somewhat different because it also incorporates syntactic information rather than portraying purely semantic classes, i.e. it treats S-, A- and P-complements separately. Moreover, it is also

---

[7] In order to calculate the proportions, I only took the first three co-occurrence categories into account (i.e. the figures from rows 1–3). For control purposes, I also conducted an analysis that included the cases of 'partial/restricted' coverage. It yielded very similar results, especially as far as the relative ranking of the predicate classes on the binding hierarchy is concerned (cf. below).

[8] Givón (1980) distinguishes three 'supra-classes' of CTPs, i.e. 'modality verbs', 'manipulative verbs' and 'cognition-utterance' verbs, but it is not difficult to extract which of our predicate classes are subsumed under each group, and how these smaller classes, in turn, are ranked in Givón's proposal.

methodologically different since it draws on a specific statistical method to derive the scale from the data: It operates with the mean CIDs of the complements of each environment to see which environments show similar CID values. Their relative similarity on this measure can then be visualized in a format that matches the idea of a hierarchy or scale. A possible way of achieving this effect is to submit the similarity measures to a technique called 'Multidimensional Scaling' (MDS; cf. Kruskal and Wish 1978 for a classic reference, and Croft and Poole 2008 for its application to typological data). Specifically, I applied metrical MDS to the data at hand and coerced the results into a one-dimensional representation, yielding the scalar format of the MDS plot in Table 20.[9] In contrast to Givón's and Cristofaro's versions of the hierarchy, this scale does not only reflect possible groupings of CTP classes, but also the actual distances between them, and it can do so because it is based on a mathematical formalization of dissimilarity relationships and their representation in geometrical space (cf. also Croft and Poole 2008: 6f. on this point). The location of individual classes on this scale will be incorporated into the following paragraphs, where many environments are discussed separately. With Table 20 as a foil, we are now in a position to do just that. For organizational purposes, I will divide the examination of the CTP classes into S-, A- and P-related environments.

6.3.2 Environments in P-function

We will begin by looking at the most familiar group of CTP classes, i.e. those that involve complements in non-subject functions. Since these are at the heart of Givón's (1980) paper and Cristofaro's (2003) investigation, no elaborate treatment of each class will be offered here. I will concentrate instead on the interplay of productivity, alternative expression formats and the relative position on the binding hierarchy.

From the perspective of productivity, the different environments in the P-domain are not created equal: If we submit the CIP values in row 8 of Table 20 to a Kruskal-Wallis analysis again, we can see that the data are significantly skewed overall ($H = 345.4$, $df = 16$, $p < 2.2e\text{-}16$). The highest values in the P-domain are achieved by jussive environments, closely followed by the perception class and all environments to the right of it in Table 20 (i.e. knowledge, propositional attitude, emotive and quotative). The fact that jussive environments slightly outrank those others may be due to their sharing semantic features of several other environments simultaneously: First, jussive predicates typically involve an utterance component ('order, tell, ask') and hence act as attractors for quotative complements. Second, they involve a manipulative component and may thus come to be coded by a causative complementation pattern (with appropriate TAM modifications). Third, jussive predicates are directed at an unrealized, potential state of affairs, which they share with desiderative complements and purposive constructions. It is this multiple overlap that makes jussive environments hospitable to many different kinds of complementation patterns and may hence result in their high cross-linguistic productivity. However, it has to be

---

[9] The relevant routines in *R* are `dist` for the calculation of a distance matrix, and `cmdscale` (with k = 1) for obtaining the MDS solution. Note that the dissimilarity calculation takes only the mean CID into account and disregards the dispersion of the data around the mean in each CTP group.

emphasized that upon closer statistical evaluation of the data, the CIP of jussive predicates does not differ significantly from the other classes mentioned above, at least not under very conservative estimates.[10]

Crucially, though, the remaining environments in the P-domain (i.e. desiderative, phasal-P and causative classes) appear to have a considerably lower CIP than the others, as the raw figures in row 8 suggest. It is telling that these four classes are precisely the ones that code Dixon's 'secondary' rather than 'primary' concepts, so the distinction seems to find some empirical support. Statistical scrutiny reveals that phasal-P and causative environments do, indeed, have a *significantly* different degree of productivity from the remaining P-environments, but the two desiderative groups turn out not to be significantly different: As the absolute figures in Table 20 show, the two desiderative environments sit somewhere in between the classes that tend to attract complement clauses cross-linguistically (i.e. jussives, perception, etc.) and those that do so significantly less often (i.e. phasals and causatives). Let us have a brief look at these 'outliers'.

Phasal-P predicates score relatively low on complement productivity because they are particularly prone to serialization and especially auxiliation, i.e. they are complement-taking predicates from a diachronic point of view but have reached the status of monoclausal complex-predicate constructions synchronically (as discussed in §2.3.1). This type of grammaticalization is propelled by the tight semantic integration of the two states of affairs in phasal constructions: beginning, resuming or finishing an action is already an integral part of the action as such, and so it is quite a natural development for the two verbs involved to fuse into a single complex predicate (cf. Cristofaro 2003: 118f., 252f. for detailed discussion). As Croft (2001: 260) shows, the process of auxiliation then involves a partial or full redistribution of the head-dependent relationship between the two verbs, so that the former matrix verb ends up being an auxiliary to the former complement verb, which itself becomes the head of a simple sentence. Examples of this process were provided in §2.3.1 above, and apart from phasal verbs, auxiliation is also common for causative and desiderative predicates. Heine and Kuteva (2002: 328) show that causative auxiliaries typically go back to CTP uses of 'give', 'do' and 'take', and my data exhibit quote a few cases in which desiderative verbs, too, are argued to be auxiliaries or uninflecting particles from a language-specific perspective. This holds, for example, for Ainu, Burmese, Hup, Newar, Trumai and others. An even more advanced stage of grammaticalization is reached where these secondary concepts morphologize and become affixes on the main verb. This, too, has been discussed at length in the literature, and the list of languages in my sample that employs morphological strategies rather than complement-taking

---

[10] This can be established by performing post-hoc pairwise comparisons on the CIP distribution as a whole. Similarly to a parametric ANOVA analysis, the non-parametric Kruskal-Wallis statistic can also be checked for the individual contrasts that contribute to the overall significance. To this end, one can compare the median CIP for all pairs of environments in the data, with a *p*-value that is appropriately adjusted to the fact that multiple comparisons are made on the same dataset. In *R*, this can be done very efficiently with the routine `kruskalmc`, implemented in the package `pgirmess`. (I thank Patrick Giraudoux for making this package available to the *R* community.) Note, though, that the output of this application only indicates which pairs are significant (at the level of α = 0.05), but without providing a precise *p*-value or effect size. Therefore, whenever such pairwise comparisons are reported on in this chapter, no significance values are given.

verbs for at least one of the above concepts is endless.[11] I content myself with providing just two examples here:

(147) Tepehua (Totonacan: Mexico; Kung 2007: 263)

Juu Xiwaan maa-k'uk'a-y   juu x-k'iw   juu x-maak'uk'aʔ.
ART Juan **CAUS**-carry-IMPF ART 3POSS-wood ART 3POSS-pack.animal
'Juan makes his pack animal carry the wood.'

(148) Evenki (Altaic, Tungusic: Russia; Nedjalkov 1997: 70)

Tyma:tne tegemi   suru-mu-d'ere-n.
tomorrow morning go.away-**DESID**-PRS-3SG
'He wants to leave tomorrow morning.'

As Cristofaro (2003: 253) notes, monoclausal strategies for phasal, same-subject desiderative and causative predicates are "possible because the linked states of affairs share their participants, and the time reference, aspect, and mood value of one of the two are predetermined." Therefore, whenever these secondary concepts *are* coded by complementation constructions rather than alternative means, the best-fitting structures from an economical and iconic point of view should be ones that leave the subject implicit (e.g. control constructions) and possibly reflect semantic integration and predetermination in other respects as well. This normally holds for desententialized, 'non-finite', patterns, and this is precisely mirrored by the data in Table 20: Phasal CTPs exhibit the highest proportion of dependent verb forms in the entire P-class (72.9%), followed by same-subject desideratives (50.6%).[12] The same ranking can also be observed on the binding scale, where the CID of phasal complements comes out as lowest in the P-domain again (as in Givón's and Cristofaro's versions). In pairwise statistical comparisons, the phasal-P environment enters into significant CID contrasts with the upper portion of my scale (i.e. with all environments from 'perception' predicates upwards). The same-subject desiderative class is the second lowest P-environment on the scale; its median CID (2.67, cf. row 12 of Table 20) is not significantly different from that of phasal predicates, but it *is* different from that of the highest members on the scale, i.e. propositional-attitude and utterance predicates.

Interestingly, what we observe on the scale is that same-subject desideratives are unlike their different-subject counterparts: The proportion of dependent verb forms is lower in the different-subject class (35% as compared to 50.6%), and the corresponding CID differences (1.85 versus 2.67) entail that the two environments are not adjacent on

---

[11] The relevant languages (e.g. Awa Pit, Gulf Arabic, Barasano, Gooniyandi, Kana, Tepehua, Evenki, Jamul Tiipay, Santali, Motuna, Martuthunira, Warao, Kayardild, Somali, Mapudungun, Tamashek, Dolakha Newar and many others) come from all macro areas. The widespread distribution of causative morphology is documented in Shibatani's (2011) *WALS* chapter. Haspelmath's (2011b) *WALS* chapter includes information on desiderative affixes and particles, and shows that these are common but appear to have a conspicuous Pacific-Rim distribution.

[12] Note that the class that closely follows same-subject-desideratives here, i.e. emotive predicates (48.9% dependent verb forms), is also particularly likely to involve same subjects (though not necessarily predetermination of the temporal, aspectual or modal value of the complement). Haspelmath (to appear) argues that human beings are egocentric and talk more frequently about desires concerning their own actions rather than other people's actions. I suspect that the same holds for emotive reactions like 'fear/afraid', 'regret', 'be happy', etc., and this bias towards same-subject expression makes emotive complements prone to be coded by subjectless, i.e. 'non-finite', structures.

the hierarchy.[13] As can be seen, Cristofaro's scale does not make a distinction between the two classes, but Givón's (1980) paper at least suggests that desiderative notions often partake in several predicate classes and that this may be related to a difference in participant sharing: same-subject 'want' predicates belong to Givón's so-called 'modality verbs', and hence are expected to pattern at least to some degree with phasals and similar verbs, while different-subject 'want' predicates constitute the lower end of Givón's 'manipulative' class. The latter often shades seamlessly into his 'cognition-utterance group', which in turn is the one with the lowest semantic binding and syntactic integration. My findings thus lend empirical support to a distinction between two different types of 'want'-scenarios (cf. also Khanina 2009), and an illustrative example of a language-internal split between them comes from Urarina. As can be seen in (149) below, same-subject desiderative clauses are coded by a nominalized construction called the 'Infinitive', while different-subject desideratives take what Olawsky (2006: 434) calls a 'finite' subordinate clause:

(149) Urarina (isolate: Peru; Olawsky 2006: 767, 769)

    a. [*Ʉna*]   *heri-to-anʉ.*
       come-**INF** want-NEG:3
       'I (really) want to come.'

    b. [*Ita*   *ajna*   *kauatɕa-ĩ*   *ni-aka=ne*]   *here.*
       RECP   with   good-PTCP   be-1PL/DU=**SUB**   want:3.EF
       'He wants us to be nice with each other.'

Despite such structural differences between same-subject and different-subject desiderative environments, their overall productivity in complementation systems is similar – and somewhat lower than that of typical complement-taking predicates, just as predicted by Dixon's idea of secondary concepts. For same-subject desideratives, it was shown that this is because of their propensity for grammaticalization (auxiliation, affixation), but this applies to different-subject desideratives only to a limited extent: Haspelmath (to appear) argues that the use of desiderative affixes is uncommon for different-subject scenarios. In my sample, it does occur, for instance, in Matsés:

(150) Matsés (Panoan: Brazil, Peru; Fleck 2006: 228)

    Mibi   *cho-paşhun.*
    2ABS   come-NPST:**DESID**:2/3
    'I want you to come.' ('(I wish/it would be great if) you could come.')

Instead of a dedicated desiderative affix, quite a few languages employ subjunctive or optative morphology in simple sentences to evoke the meaning associated with different-subject desideratives. Examples of this pattern can be found in Nkore-Kiga or

---

[13] The differences are not significant under the most conservative estimates (i.e. non-parametric pairwise CID and verb-form comparison with *p*-value adjustments), but if we directly compare the proportion of dependent verb forms without adjusting the *p*-value, a marginally significant result arises (Fisher exact test, *p* = 0.057, odds ratio = 1.89). However, Haspelmath (to appear) shows that the morphosyntactic differences between same-subject and different-subject desideratives are more subtle than the effects captured by the verb form and the CID here (e.g. differential complementizer omission, shortness of the verb 'want' as such, etc.). Therefore, despite the fact that Haspelmath's sample is not controlled, there is reason to believe that the difference between same-subject and different-subject desideratives is actually profound.

in the Australian languages Gooniyandi, Kayardild and Wambaya. Thus for the latter, Rachel Nordlinger (p.c.) reports that different-subject desiderative meanings "are usually done without the 'want' verb: 'she should go'. I never came across a 'want' verb in these contexts." Such languages, then, have simply not grammaticalized a different-subject 'want' environment in complementation but render this meaning by other techniques. Sometimes it can only be paraphrased. For Tümpisa Shoshone, Dayley (1989: 385) reports that "there is no direct equivalent of [the different-subject desiderative pattern]; the closest would be with a verb of telling instead of wanting." For Awa Pit, Curnow (1997: 166) observes that "when a situation is to be described where one person wishes another to carry out an action, a periphrastic way of stating this is used, usually something like 'it would be good (for me) if you …'". Finally, it is also possible for the different-subject-want environment to be rendered syntactically as a subject clause rather than an object clause (lit. 'It wants me that he go', as in Yuracaré; more on such patterns below). In these cases, too, the specific environment we are looking for is not attested in the complementation system and represents a 'constructional gap'. All of the alternative strategies lower the productivity of different-subject desideratives as selectors of complement clauses and may thus help explain the data in Table 20.

Causative CTPs are interesting for a similar reason: Being a secondary concept in Dixon's framework, their productivity in complementation is severely restricted; in fact, they show the lowest productivity figures of all P-environments in Table 20. As was noted above, the primary alternative for causative complements is affixation ('I go-made him'). But when causatives *are* rendered by complementation, their preferred structural choice in my data is not that of desententialized structures: The proportion of independent verb forms is 68%, and their mean CID is such that it fits into the upper portion of the binding scale. It is here that my hierarchy differs from those of Givón and Cristofaro, both of which list causatives immediately after phasal predicates and thus with a high degree of desententialization. In Givón's case, this is simply due to the fact that his hierarchy is explicitly supposed to reflect that causatives tend towards clause union, lexicalization ('lexical causatives') or grammaticalization (causative affixes); in other words, it is a hierarchy that comprises more grammatical techniques than complementation alone. And the difference to Cristofaro's hierarchy may be grounded in the fact that she deliberately neglected predicates of more indirect causation, such as 'cause' itself, while I adopted a more embracing approach. In other words, while I did split the supra-group of 'manipulative' predicates (e.g. Noonan 2007) into causative and jussive environments, I did not make any internal gradation as to the strength or immediacy of the causation ('make' versus 'force' versus 'cause'), and this may explain why causative complements appear in a different place on my binding scale.

Let us finally turn to those P-classes that code primary concepts in Dixon's sense. As was mentioned above, all of them are highly productive (without any significant differences between them), and their relative position on the binding hierarchy conforms to previous research. In Givón (1980), knowledge, propositional-attitude and utterance predicates are included in a larger rubric of 'cognition-utterance verbs'. These are argued to be internally graded according to certain semantic principles

which he subsumes under his notion of 'binding' (here: the degree of emotional commitment of the matrix agent to the truth or realization of the complement event). This binding difference yields the cline 'hope/remember/forget' > 'think/believe/know' > 'say'. In terms of the raw figures (verb forms, CIDs), my data can confirm this, but in keeping with Cristofaro (2003), the group-internal differences between the relevant notions (e.g. 'think/believe', 'know', 'say') are not statistically significant, so that a genuine ranking is infeasible here. This empirically reflects Givón's suspicion that "languages do not always exhibit many coding points along [the] cognition-utterance verb portion" (Givón 1980: 363), or these are too subtle to be captured by the constructional distinctions in the present study.

What is presumably also included in Givón's 'cognition-utterance' group is the class of perception verbs; Givón does not discuss them explicitly. However, if they are, then it has to be pointed out that they show markedly different structural preferences from the other cognition-utterance verbs. Ransom (1986: 26) already argued that perception verbs show this anomalous behaviour (and hence criticized Givón's general account), and Cristofaro (2003) supports this empirically. My data on perception verbs are fully consistent with these latter findings. As can be seen in my version of the binding scale, there is a gradual increase in the mean CID from utterance over propositional-attitude, knowledge, different-subject desiderative, causative and jussive to perception contexts. The transition is gradual, but the perception class is the first one that is *significantly* different from the top of the hierarchy, i.e. it enters into a significant CID contrast with utterance verbs. This is also supported by the data on verb forms, where utterance and perception verbs differ significantly in their proportion of dependent forms (17.3 versus 40.2%, Fisher exact test, $p = 0.00038$ ($< p_{adjusted} = 0.003$), odds ratio = 3.18). I mentioned earlier that perception verbs are often compatible with several different types of complement in a given language, but the situation is typically such that the more sentential type of complement yields an evidential reading, while the desententialized one is reserved for immediate perception. Since we are exclusively concerned with the latter type of perception, it follows that such situations lead to an overall bias of my 'perception class' towards desententialized complements. In the literature – most comprehensively by Cristofaro 2003 –, a number of reasons have been discussed for the morphosyntactic differences between the complements of typical cognition-utterance verbs on the one hand and those of perception verbs on the other. They chiefly pertain to the semantics of perception verbs (cf. also Kirsner and Thompson 1976 for a classic study), and include (i) that only perception verbs predetermine the TAM configuration of the complement (= simultaneity of the two situations, cf. also Ransom (1986) on this point); (ii) that perception complements relate to the matrix clause on the predicational rather than the propositional level (cf. Dik and Hengeveld 1991); (iii) that there is some degree of semantic integration and interlacing in perception complements: the perception of an entire situation always entails the perception of the participants involved in this situation, hence perception complements are open to structures that code one of the perceived participants as the object of the perception verb (object control or raising constructions, as discussed in detail in Chapter 4). Taken together, this specific constellation of properties may explain why environments of immediate perception can be 'targeted' diachronically by

complementation structures that are different from those of fully propositional cognition-utterance complements (cf. also Horie 1993). A nice illustration of this comes from Burmese, for which Soe (1999: 304–5) reports:

> Finite sentences may be directly embedded as quote complements of utterance verbs. They […] occur as complements of cognition verbs, and may do so only when the cognition verb allows a spatio-temporal separation between the act of cognition and the event coded by the complement. [...] On the other hand, finite sentential complements do not occur with perception verbs like *myin* 'see', *ca:* 'hear' etc. where the act of perception and the event perceived have to be co-temporal.

This is illustrated in the following examples, where a fully sentential type of complement in (151a) contrasts with the clausal nominalization employed for perception contexts, as in (151b):

(151) Burmese (Sino-Tibetan, Tibeto-Burman, Burmese-Lolo: Myanmar; Soe 1999: 304–5)

    a. [*Thu lou' hcin <u>te</u>*] *htin* <u>*te*</u>.
        3   do want RL   think RL
        'I think that she wants to do (it).'

    b. [*Hka-lei: <u>t</u>wei hso. nei <u>ht</u>a   <u>k</u>ou*] *ci.* *<u>h</u>cin* <u>*te*</u>.
        child   PL   play stay **NMLZ** OBJ   see want RL
        'I want to see the children playing.'

As can be seen, the structural bifurcation in Burmese is such that utterance and propositional-attitude verbs draw on a fully sentential structure (possibly quotative in origin, because the complement can optionally be enhanced by a form related to the manner particle *lou*, cf. Güldemann 2008 for manner deixis in the quotative domain); perception complements, by contrast, employ a process of clausal nominalization, where the complement is firmly embedded as the object of 'see' (cf. the object marker at the end of the complement). Interestingly, this latter structure is ultimately derived from a relative clause (by fusion of the realis relativizer <u>*te*</u>. and the head noun *ha* 'thing' into the clausal nominalizer <u>*hta*</u>, cf. Soe 1999: 299). More generally, the synchronic overlap between relative clauses and perception complements is conspicuous, and we will return to this issue in Chapter 7.

As was stated at the beginning of the present section, a lot more could be said about how the 'binding issues' in P-environments play out in individual languages, and how they need to be explained. But given the extensive amount of discussion of these points in the previous literature, I will not dwell on this here. Let me instead conclude the section by pointing out that even the more productive P-environments, i.e. those encoding primary concepts, can, of course, exhibit constructional gaps and be rendered by alternative means instead. This is obvious for quotative environments, which can be expressed by the loose concatenation of independent discourse units rather than in a subordinate or embedded type of relationship, and grammaticalization may even lead to monoclausal structures that involve a quotative particle rather than an utterance predicate. The latter has been reported, for example, for direct speech in Semelai and Jamul Tiipay (cf. also Spronck 2012 for discussion of this pattern). Monoclausal strategies involving particles may also be found in the propositional-attitude domain.

For Mangarayi, Merlan (1982: 4) relates that "there is no Mangarayi verb comparable in sense to English 'think' which represents a speaker's attitude. Instead, there are a number of mode particles [….]. There is furthermore no verb which can be used to express that what is being represented is thought rather than speech." In Yuracaré, 'think' is taken care of by Assumptive and related inflections in the verb morphology (van Gijn 2006: 243) and, likewise, does not exist as a CTP. In Hup, evidential morphology stands in for perception verbs: Simple sentences may thus occur in the unmarked evidential form indicating visual evidence, or in a marked form signifying 'nonvisual but still first-hand' evidence (Epps 2008: 643, 646). And finally, an interesting biclausal alternative that I encountered for several P-classes is a chaining construction in which the 'CTP' occurs as the *dependent* ('medial') verb; because of this reversal of the 'complement-taking relationship', these structures are best seen as periphrastic devices rather than complement clauses or genuine complementation strategies. In Motuna, for example, causative relationships are most productively rendered by lexical or morphological causatives, but there is a biclausal alternative in which predicates like 'persuade/force/make' appear as medial verbs in a clause chain; being a looser type of clause combination, this technique implies a certain degree of control being left to the causee:

(152) Motuna (East Bougainville: PNG; Onishi 2000: 127)

[*Kongot-us-uu*] *pi-i-ng.*
persuade-3P<1A-CONT.DS go-3S-NR.PST-M
'While I persuaded him, he went.' (= 'I forced him to go.')

Exactly the same structure is found for jussive environments in Lavukaleve ('he-telling-them, they went' = 'He told them to go', cf. Terrill 2003: 424), and as a periphrasis of perception verbs (in addition to a nominalized complement) in Kolyma Yukaghir ('my seeing him, he went' = 'I saw him leave', cf. Maslova 2003: 391). These examples will suffice to convey the general idea.

6.3.3 Environments in S-function

We can now proceed to the environments that have not yet received much attention in the literature, viz. those in which the complement appears in subject function. The present section is devoted to monovalent main clauses with a complement in 'S' function. Examples of such S-clauses have already been given throughout the preceding chapters, especially on CTPs like 'good', 'easy', 'seem' or 'important'. What is perhaps less evident is the semantic diversity hidden in the group called 'other-S' above. We will thus start by illustrating some of the concepts that have been found lexicalized as CTPs taking an 'other-S' clause in my sample:

(153) Malayalam (Dravidian: India; Asher and Kumari 1997: 42)

[*Avan varum ennatə*] *aaʃcaryyakaram aaŋə.*
3SG.M come-FUT QUOT-NMLZ surprising be-PRS
'It's surprising that he will come.'

(154) Dolakha Newar (Sino-Tibetan, Tibeto-Burman, Bodic: Nepal; Genetti 2006: 144)

[*Thi-mā rājā nāplaŋ-a*] *ju-en* *con-a.*
one-CLF king meet-NMLZ become-PTCP stay-3SG.PST

'It turns out they met a king.'

(155) Jamsay (Niger-Congo, Dogon: Mali; Heath 2008: 603)

[*Dè:né jìné   jǎ: àbádá kúdáy   yɔ́=kɔ̀*]   *kárⁿá   bɛ̀rɛ̀-gó-Ø.*
 keep hold take always for.good exist=be.NHUM   be.done can-IPFV.NEG-3SG.SBJ
'Keeping and storing (millet, so) there is always some (at hand), cannot be done.'

(156) Mayogo (Niger-Congo, Adamawa-Ubangian: DR of Congo; Sawka 2001: 81)

Má gala kolo   [*me   kpʉlʉ a-koto*].
like quick PST-arrive COMP snake PST-appear
'Quickly (it) arrived that (a) snake appeared.'

(157) Musqueam (Salishan, Central Salish: Canada; Suttles 2004: 522)

Háy   niʔ   x̌ʷə́m [*kʷ-s-[ʔi]-s*   *ʔəⁿi   wəq̓ʷíləm*].
specify AUX fast   ART-NMLZ-AUX-3POSS   come downstream
'Coming downstream was faster.'

(158) Motuna (East Bougainville: PNG; Onishi 1994: 489)

[*Hoo   roko-ng   haa-warei*] *haha'w-oro-kuu-ng.*
 ART.M you.SG.POSS-M want-NMLZ work-MID.3S-IMAG-M
'Your wanting [= wishes] could have worked.'

(159) Akkadian (†Afro-Asiatic, Semitic: Mesopotamia; Deutscher 2000: 54)

[*Kīma tupšikk-um   šū   lā   labirta-šunu*]   *innamer.*
 COMP corvee.work-NOM DEM:M.SG:NOM not old.debt-3M.PL.POSS   3M.SG.saw.PASS
'It was seen that his corvee work was not their old debt.'

Examples (153) and (154) occur in contexts of surprise; this is a common implicature in (154) (cf. Genetti 2006: 144) and lexicalized as a CTP 'be surprising' in (153). As an alternative to such CTPs, languages may employ specific verbal morphology that is distinct from evidentiality contrasts; DeLancey (1997) thus proposed the grammatical category of 'mirativity' for such dedicated 'surprisal' morphemes. The specific context in (155) shades into abilitative or deontic modality (depending on the context), but Jamsay also has a more general negative predicate '(be) not' that can be complemented in the same way, i.e. where Heath (2008: 603) concludes that "there is no reason not to assume that the [clause] functions as subject." In fact, Dixon (2006a: 12) lists negative predicates 'not, don't' among the secondary concepts that can take an S-clause. He mentions Fijian as a language with such a lexicalization pattern, and in my sample, I have also found it in Hup (Epps 2008: 850). In (156) from Mayogo, we see the predicate 'arrive' in the sense of 'happen'; this is what Ransom (1986: 37) calls an 'occurrence modality' complement (which in English is also possible with CTPs *take place, come to pass, develop*, etc.). And apart from the mere occurrence of an action, some part of its temporal-aspectual contour or its manner may be referred to by the CTP; this can be seen in (157) from Musqueam, and further examples include CTPs like 'be close' (e.g. Kwazá), 'last long' (e.g. Musqueam again) and 'be common' (e.g. Turkish). Finally, example (159) from Akkadian (which was cited in Chapter 5) is included here again because it illustrates another context for the occurrence of S-clause, namely that of passivized transitive CTPs. Van Valin and LaPolla (1997: 268) refer to the subjects of passive clauses as 'derived-S' arguments, and such arguments can, of course, also take the form of a clause. What is noteworthy about Akkadian is

that such derived-S contexts are the only possibilities for a finite *kīma*-complement to assume the syntactic function of subject (Deutscher 2006: 162), i.e. S-clauses are not productive with other CTPs.

This fact from Akkadian leads us to a more general distributional characteristic of complement clauses: While all sample languages have complements in P-environments, the same is not true for subject clauses, neither in S nor in A function. Thus some authors have been unable to identify such constructions in their respective languages. For Mosetén, for instance, Sakel (2004: 429) reports that there are no examples of subject clauses in the corpus, and her attempt at elicitation did not yield a single unequivocal example of what one may call a 'subject clause' in the present sense. Focusing on S-clauses for the moment, a total of 15 languages in my sample do not feature this construction type in the materials available to me, and except for Krongo and Kiowa, all of them are found in either Australia-New Guinea (Abun, Imonda, Kayardild, Kewa, Mangarayi, Martuthunira, Skou, Wambaya, Warembori) or South America (Mekens, Mosetén, Trumai, Warao). For other languages, it has been reported that S-clauses are possible but "much more limited" than object clauses (e.g. Epps 2008: 850 on Hup). Similarly, Weber (1994) states on Huallaga Quechua that S-clauses are rare in general, and that, depending on the specific CTP, they may be understandable but are difficult to elicit.[14] A recurrent pattern, then, appears to be a restriction on the productivity of S- (and A-) clauses as compared to P-clauses.

However, while the data in Table 20 generally confirm this, it can also be seen that the 'evaluative-S' environment achieves the highest productivity score (i.e. CIP) of *all* CTP classes in the present study! This may be attributable to two facts: First, the environment itself is not particularly constrained: it comprises the evaluation of a proposition ('It is good that you've resigned') as well as an evaluation relating to the performance of an act, either generically or in regard to a specific instance ('Passing this test is easy', 'It was easy for him to pass this test'). Consequently, the environment does not impose any tight constraints on the structure of the complement, which in turn makes it hospitable to a variety of different structures (where these exist). Second, the evaluative-S environment chiefly consists of primary concepts in Dixon's sense, many of which are lexicalized as predicates in the world's languages because they are communicatively very useful: "Valuations constitute an aspect of all categorization [processes]", and according to many researchers, there is a "human urge to externalize values verbally", so that "the expression of values is an all-pervading feature of language." And the "lexical level", in turn, "is the most evident level at which we can see evaluation at work" (Alba-Juez and Thompson 2014: 4–9). It is thus no surprise that many languages have predicative adjectives or stative verbs with evaluative denotations. This holds for simple sentences, where a quality is evaluatively predicated of a referent ('This joke was good'), and for complex sentences, where a quality is

---

[14] For example, S-clauses on the evaluative predicates '(not) good' "fue bastante difícil elicitar. […] Ninguno de [hablantes] tuvo dificultad para entenderlo, pero todos dijeron que era una construccíon rara." (Weber 1994: 101) In other words, elicitation of contexts involving '(not) good' and a complement clause proved rather difficult; while the informants had no difficulty understanding such structures, they agreed that it is a rare construction type. Interestingly, many of the adjectival predicates that are felt to be more acceptable with an S-clause are borrowed from Spanish, so there may actually be some structural calquing involved.

evaluatively predicated of a proposition or a state of affairs ('Playing fair is good'). As a result, the evaluative-S environment in Table 20 outranks all other S-taking contexts in complementation, and its CIP enters into an individually significant contrast with all of them.

Apparently, then, the other environments in S-function are prone to alternative techniques of grammatical coding. At first sight, their relative productivity in S-complementation seems to correspond very neatly to their instantiating a primary or a secondary concept in Dixon's sense. For example, the 'seem-S' class, which Dixon (2005: 202ff.) discusses at length as a clearly secondary concept, has a comparatively low productivity as an S-taking environment. Although it was beyond the scope of the present study to look for all alternative coding strategies for the relevant concepts (since my primary concern was, of course, with those cases that *are* lexicalized as CTPs), some such techniques were readily identified. In the case of 'seem', an alternative to a CTP rests in evidential morphology or particles. As was mentioned in §2.5.2, 'seem' is used when an inference is made based on some sort of perceptual evidence (e.g. seeing the result of an action without necessarily having witnessed the action itself) or hearsay. This is realized, for instance, by the Inferred evidential *cud* in Hup, and by the Apparential affix in Kwazá ("deductive evidentiality", van der Voort 2004: 420):

(160) Hup (Vaupés-Japurá: Brazil, Colombia; Epps 2008: 651)

*B'ŏy   yo-hipãh-nɨ́h=**cud**   ʔám-ã́h.*
traira.fish   dangle-know-NEG=**INFR**   2SG-DECL
'It seems/looks like you don't know how to carry traira fish.'

(161) Kwazá (isolate: Brazil, Colombia; van der Voort 2004: 421)

*Maga'riDa   o'nɛ-**tehere**.*
Margarida   fall-**APPAR**
'Margarida seems to have fallen.'

Similarly, languages like Wappo or Wari' employ an evidential particle inserted into a simple sentence, but no CTP 'seem'. The diachronic sources of these markers are quite diverse (cf. Aikhenvald 2004: §9.2), and as with other environments of complementation with lower productivity (e.g. phasal-P, causatives, etc.), there is some suggestive evidence that deductive evidentials may actually be grammaticalized forms of former complementation patterns on a CTP 'seem' or a similar predicate. In Kana, for instance, the comparative lexeme *bēē* 'resemble' has fused with the complementizer *kɔ̀ɔ̀* to constitute a "formative *béékɔ̀ɔ̀* […] in sentence-initial positions" (Ikoro 1996: 352):

(162) Kana (Niger-Congo, Benue-Congo, Cross River: Nigeria; Ikoro 1996: 352)

Béékɔ̀ɔ̀   nɛ̄ɛ̄   a̱a̱   lū.
resemble+CONN   person   PROG   come
'It seems that someone is coming.'

Similarly, in Turkish, a lexical verb 'seem' has combined with the postposition and adverbial conjunction *gibi* 'like' to form a "verbal idiom" (Göksel and Kerslake 2005: 412), which can now occur again as a CTP with an unmarked complement clause (or a simple sentence, depending on the analysis one wishes to assume). Such gravitation

towards monoclausal techniques would sit well with Dixon's idea of secondary concepts.

But following Dixon's line of argumentation is not always as straightforward as for 'seem'. This is because Dixon (2005, 2006a) himself does not discuss all relevant classes in regard to his primary/secondary distinction. He is explicit about 'seem', phasals and 'not' being instances of secondary concepts, but the remaining ones must be inferred from his discussion. To the extent that his notion 'must' is a secondary concept, predicates like 'be necessary' with a deontic-S complement may also be seen as instantiating a secondary concept. In Dixon (2005: 282), we further find a statement that adjectival predicates from our epistemic group ('true', 'certain', 'unlikely') "are most different in meaning from verbs", which I take to mean that they – just like the conceptually related 'seem' – qualify as secondary concepts modifying primary ones. This contrasts with most predicates in the evaluative group ('good/bad', though not necessarily 'easy/difficult'), which Dixon (2005: 279ff.) subsumes under complementation with primary verbs.[15] 'Other-S' predicates, finally, would be mixed since they comprise both secondary notions (e.g. 'not') and primary ones (e.g. 'happen' (Dixon 2005: 97) or the passive counterparts of primary concepts like 'see', 'hear', 'order', etc.). Again, the whole distinction between the different concept types may not be fully convincing to begin with, but if we accept it as a proposal made in the typological literature on complementation, then it would appear that the highest amount of primary types in relation to S-clauses is found in the evaluative-S class, followed by the other-S class, while the remaining classes are clearer instances of secondary concepts. This is the coding shown in Table 20 above (cf. row 9). If this classification is adopted, the distinction between primary and secondary concepts across the subject- and object-clause domain correlates with a significant difference in productivity as a complement-taking environment. Indeed, the mean CIP for classes encoding alleged primary concepts is significantly higher than that of secondary-concept classes ($mean_{primary} = 95.8$ ($SD = 18.1$), $mean_{secondary} = 51.8$ ($SD = 25.3$), $t = 4.0$, $df = 12.53$, $p = 0.001$), and this also holds under a non-parametric comparison ($median_{primary} = 99$ ($IQR = 9$), $median_{secondary} = 55.5$ ($IQR = 24.25$), $W = 66$, $p = 0.004$)). This is visualized in the following boxplot (Fig. 13):
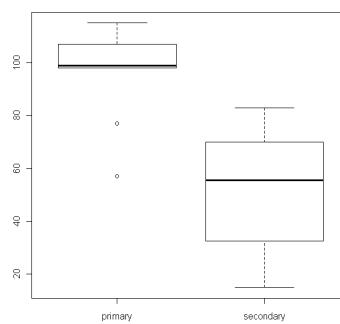
Figure 13. Average CIP scores for primary and secondary concepts according to Dixon (2005, 2006a)

---

[15] Note that the contributions to Dixon and Aikhenvald's (2006) volume on complementation treat semantically similar evaluative predicates sometimes as primary, sometimes as secondary notions, and it is not clear if this follows the same principles in each case.

The two outliers among the primary concepts are A-clauses, whose low productivity will be discussed in §6.3.4 below, and 'other-S' clauses, which we have just shown to be actually mixed between primary and secondary notions; its mean CIP is thus in perfect keeping with this.

Despite these results, there are several complications to the seemingly neat picture above. The first is, of course, that one may generally object to Dixon's classification because no cogent independent reasons for assuming such a conceptual distinction are provided. It could be asked, for example, why 'want', 'avoid', 'help' and 'permit' are *categorically* different from primary concepts like 'think' or 'prefer' (cf. Dixon 2005: 97f.). Moreover, the primary concepts that necessarily co-occur with a secondary one (e.g. *I helped my son to <u>tie</u> his shoelaces, She avoided <u>going</u> through the park,* etc.) are said to be "central" in their respective clause "from a semantic point of view" (ibid.), even if they occur syntactically as CTPs governing a complement clause (as in the examples given). But this assumption, too, need not be shared by everyone. (Langacker (1991: 436), for one, proposes that complementation structures like *I helped my son to tie his shoelaces* profile the act of helping, not that of tying the shoelaces.) In light of these reservations, one may say that Dixon's distinction is intuitively plausible, and appears to be supported by the above data, but it remains somewhat arbitrary when applied to specific concepts and linguistic expressions.

Second, the values in Fig. 13 aggregate the data from subject and object clauses and hence conceal the fact that even within the group of the alleged secondary concepts, there are significant productivity differences depending on whether the concept is lexicalized in the subject- or the object-taking domain. This applies to the phasal domain, where phasal-P environments (*He began to work*) are significantly more productive in complementation than phasal-S (*It began that he worked*). That is, regardless of the fact that phasal-P contexts are themselves extremely prone to grammaticalization, they are still more productive as genuine CTPs than phasal-S constructions. A similar contrast between semantically equivalent classes (although they are not both secondary concepts in Dixon's framework) is found in the domain of propositional attitude: Here, too, P-contexts like *I think/believe …* have a much higher average CIP than the corresponding S-contexts *It is likely/certain that…* Taken together, there appears to be an additional principle at work that cuts across the primary/secondary-concept distinction: Personal construals of events are generally preferred to impersonal ones in the lexicalization of CTPs. In other words, CTPs tend to be lexicalized in such a way that they themselves predicate something about a human participant and thus construe the whole complementation construction from the perspective of this participant. We will see this most clearly in the upcoming section on A-clauses, but there is evidence for the same principle in S-environments, too. This evidence relates to the fact that, across all of the S-taking environments in Table 20, an alternative coding strategy can be found that has not yet been discussed: In this strategy, the concept is not rendered by affixes or particles, but remains a CTP (or an auxiliated version of a former CTP); crucially, though, the CTP is lexicalized with a personal subject and the complement clause as its object. Let me provide a few illustrative examples from the relevant domains.

With regard to 'seem', it is well-known from English that beside an impersonal construal (*It seems that John is ill*), a personal one is also available (*John seems to be ill*, cf. also the translation of example (161) above). In the literature, the latter has often been described as involving a 'raised' variant of the former, but as we saw in §4.4.1, languages may also lack the impersonal version altogether. We encountered the case of such 'obligatory raising' in Barasano in Chapter 4, and it is attested in a similar way for Mosetén. In this language, the predicate *se'we'* 'seem' is described as one of those "transitive matrix verbs" that take "an object complement clause" cross-referenced by the third-person feminine object index on the CTP (Sakel 2004: 430):

(163) Mosetén (Mosetenan: Bolivia; Sakel 2004: 430)

> *Aj* [*pho-ki-'*   *katyi'*]   *se'w-e-'-in.*
> yet   smell-V-F.SBJ   HRS   seem-V-**3F.OBJ**-PL
> 'It seems to them that something is smelling.'

Similar constructions are also found in the deontic domain. Apart from the fact that the productivity of this domain is generally limited by the pervasiveness of morphological encoding (obligative or similar affixes), many languages resort to personal construals of obligation: The CTP takes the experiencer as its subject, and a complement clause as object. This is illustrated in (164) below from Wappo, for which Thompson et al. (2006: 77) "have not discovered any alternative ways of expressing deontic modality":

(164) Wappo (Wappo-Yukian: USA; Thompson et al. 2006: 77)

> *Cephi*   [*šawo*   *ca*   *pa?-ukh*]   *ne?- khi?.*
> 3SG.NOM   bread   plain   eat-INF   have-STAT
> 'He has to eat plain bread.'

Similar patterns, partly involving the auxiliation of the former CTP, are common in African languages (e.g. Kana, Noon, Nkore-Kiga, Supyire), but are also attested in Vietnamese, Moseténan or Tetun; the latter has a number of deontic auxiliaries which are used personally rather than impersonally (although some of them may go back to an impersonal construction, cf. van Klinken (1999: 225, fn.)).

Various other S-domains or concepts can be encountered in personal construals. In Kwazá, the epistemic adverb 'true' can be used as a CTP, and the resulting structure involves 'obligatory raising' of the complement subject, effectively creating an object clause rather than an impersonal subject clause (cf. (165)); in Tetun, the same phenomenon occurs with the CTP 'happen' (cf. (166)); and in Choctaw, finally, a careful argument-structural analysis of the CTP 'difficult' reveals that even an evaluative CTP can occur in a personal construction (cf. (167)):

(165) Kwazá (isolate: Brazil, Colombia; van der Voort 2004: 700)

> *E'tay*   *watxi-'ki=tse*   [*jere'xwa kurakura-'wã*   *'tsje-a-nãi*].
> woman   true-DECL-DECL   jaguar   chicken-AN:OBJ   grab-1PL-NMLZ
> 'The woman believes that the jaguar killed the chicken.'

(166) Tetun (Austronesian, Central Malayo-Polynesian: East Timor; van Klinken 1999: 292)

> *Bei Lafaek*   *la*   *dadi*   [*n-á*   *Bei Kancil*].
> Mr   crocodile not   happen   3SG-eat   Mr   mouse.deer
> 'Mr Crocodile didn't happen/get to eat Mr Mouse Deer.'

(167) Choctaw (Muskogean: USA; Broadwell 2006: 274)

Charles-at *im-i̱kallo-h* [*abíika-cha*].
Charles-NOM III-difficult-TNS sick:LGR-COMP:SS
'It's hard on Charles that he got sick.'

In the last example, the intransitive CTP 'difficult' shows class-III ('dative') agreement of an experiencer, so indexation points towards an experiencer-object/oblique pattern (cf. Chapter 2) with a subject clause as the stimulus. However, the experiencer is also expressed as an overt NP and its position and case marking unambiguously identify it as the subject of the clause. Therefore, the complement clause at the end of the sentence cannot be a subject clause, but must be an object clause (in our broad sense of the term object; since the CTP is intransitive, it cannot be a direct-object clause). This analysis is also compatible with the fact that a same-subject complementizer is chosen, indicating that 'Charles' is the subject of both the matrix and the complement clause. This marking would be difficult to reconcile with a subject-clause analysis. Therefore, while the most natural translation of (167) into English takes the form of an S-clause, the original is a personal construction.

The general point, then, should have become clear. However, it must not go unmentioned that some languages in my sample do not generally abide by this cross-linguistic preference for personal structures in complementation. In these languages, CTPs are often lexicalized or grammatically construed in an impersonal way, so that subject clauses gain a typologically unusual prominence. A case in point is Karo Batak. This language has a wide range of CTPs, both in object and in subject function, so there is no particular restriction on subject clauses to begin with. But moreover, in performance, "the most common type of complementation involves a lower clause functioning directly as the subject of a higher stative or intransitive clause" (Woollams 1996: 300). One reason for this is that Karo Batak frequently employs passive (or Undergoer) constructions, and this carries over to CTPs in complementation: One often finds the matrix verbs being passivized, so that the complement clause comes to function as a 'derived-S/A' argument:[16]

(168) Karo Batak (Austronesian, Western-Malayo Polynesian, Sundic: Indonesia; Woollams 1996: 312)

Idahna [*lit kalak* *bayangen telu kalak*].
PASS.see.he be person in.stocks three person
'He saw (lit. it was seen by him) that there were three men imprisoned in stocks.'

In some other languages, the prominence of S-clauses is not tied to voice alternations, but due to the fact that impersonal construals are common in the lexicalization of CTPs. A cluster of such languages is found in the Amazonian area and comprises

---

[16] An alternative analysis, being consistent with many other Austronesian languages, would be to assume the existence of two voice systems on an equal footing, rather than an active/passive distinction. On this alternative analysis, sentences like (168) above contain two core arguments, rather than being detransitivized, and one would have to assume that the complement clause in such 'Undergoer' constructions is an A- rather than a derived-S argument. This is in keeping with the personal index 'he' on the CTP in (168) above. In my sample, another such voice system is found in Begak Ida'an (Gooudsward 2005: 338). However, regardless of the analysis adopted, the fact remains that the complement in question is a subject clause rather than an object clause, and the frequent usage of matrix-undergoer constructions makes subject clauses more prominent in these languages than in many others.

Jarawara, Wari' and Yuracaré. For Jarawara, Dixon (2006b: 110) reports that "the majority of complement clauses in the corpus are S argument to an intransitive verb". Intransitive lexicalization is common in all domains; even motion verbs are construed as S-clause-taking predicates ('We went along shooting arrows' → '[Our shooting arrows] went along'). Similarly, phasal environments take S- rather than P-clauses:

(169) Jarawara (Arauan: Brazil; Dixon 2006b: 107)

[*Oko*   *tafi*]   *hawa*   *to-ha-ke.*
1SG.POSS   eat.COMP   be.finished   away-AUX-DECL.F
'My eating is finished.' or: 'I have finished eating.'

Likewise, in Wari', a finite complement clause is of quite low productivity with our object-taking classes, but does crop up with a variety of different intransitive CTPs, functioning as its subject argument. An example is given in (170) below:

(170) Wari' (Chapacura-Wanham: Brazil; Everett and Kern 1997: 82)

Mija   *na*   [*ca*   *cao' nuncun*   *Orowao*].
much   3SG.RL:PST/PRS   INFL.RL:PST/PRS   eat   POSS.3SG:M   Orowao.M
'Orowao eats a lot.' (lit: 'Orowao's eating is a lot.')

In Yuracaré, the prominence of subject clauses is largely due to clausal A-arguments, so we will postpone their illustration to the next section. At present, I have no convincing linguistic or cultural explanation for why these languages lexicalize subject-clause-taking CTPs with such alacrity. To be sure, we probably have to reckon with areal influence here which leads to the local spread of such patterns, but it is an open question which structural or sociocultural conditions favour their emergence in the first place.

After having discussed the cross-linguistic productivity of the various S-taking environments, let us finally turn to their preferred structural choices. As can be seen on the binding hierarchy in Table 20, S-clauses appear from the middle of the scale downwards. At the very bottom of the hierarchy, we find the 'phasal-S' class, which thus exhibits the largest amount of dependent verb forms (73%) and the highest mean CID (3.9) of *all* environments in the present study (cf. (169) again from above for an illustrative example). The remaining S-classes then stretch upwards on the scale, with 'seem-S' showing the smallest amount of dependent verb forms (33%) in the S-domain, and also the lowest mean CID (2.19). The conceptual difference between these poles is rather straightforward: phasal environments target the predicational level of clause structure (relating to a state of affairs), while 'seem'-S environments target the propositional one, and without any semantic integration of the complement and the matrix event (i.e. the event of inference encoded by 'seem' is completely independent from the event encoded by the complement, which is not the case in phasal complements, as we saw earlier). These differences, then, appear to be reflected in their preferred structural choices. However, they should not be overinterpreted, for two reasons: First, under statistical scrutiny, a non-parametric comparison of the median CIDs of seem-S and phasal-S environments does not yield a significant signal.[17] In

---

[17] This was established again by applying post-hoc pairwise comparisons to the Kruskal-Wallis test for the CID distribution as a whole.

other words, the S-taking domain may be structurally graded in absolute terms, but it is still slightly too homogeneous to show significant internal cut-off points.[18] Second, despite the fact that the 'seem-S' class codes propositional (rather than predicational) complements, its position on the binding hierarchy is still different from that of the P-classes coding such propositional complements. As can be seen, the mean CID of the 'seem-S' environment is even higher than that of the perception-P class, which we already know to differ significantly from the top of the hierarchy. Therefore, it seems to be the case that S-clauses, as a whole or on average, tend more strongly towards desentencialization than P-complements. We will return to this point after having examined A-clauses, which are next on our agenda.

6.3.4 Environments in A-function

In §2.5, we defined A-clauses as encoding the stimulus in a two-argument clause, which induces a change of mental or physical state an in animate experiencer. The relevant predicates thus fall under the 'experiencer-object' lexicalization pattern established in Chapter 2. As with S-clauses, we shall begin by providing some illustrative examples of the range of A-complement clauses from my data:

(171) Purépecha (Tarascan: Mexico; Chamereau 2000: 229)

['*INki* *i'ma* *xu'ɽa-ka*] '*sesi* '*¢ipi-šiN-ka-rini.*
 COMP DEM come-SBJV well please-HAB-ASS.1/2-1OBJ
'That this one is coming pleases me.'

(172) Vietnamese (Austroasiatic, Mon-Khmer, Viet-Muong: Vietnam: Nguyên 1997: 222)

[*Thẳng cháu* *Hải nói* *líu-lo*] *khiến các bác* *buồn cười.*
 boy nephew Hai speak babble cause PL uncle feel laugh
'Little Hai's babbling made all uncles and aunts smile.'

(173) Amele (Trans-New Guinea, Madang: PNG; John Roberts (p.c.))

[*Hina ene* *h-og-a* *eu*] *filicit-t-ena.*
 2SG here come-2SG.SBJ-HOD COMP startle-1SG.DO-3SG.SBJ.PRS
'That you came here startles me.'

(174) Basque (isolate: France, Spain; Hualde and Ortiz de Urbina 2003: 202)

[*Zu* *hemen ikus-te-a-k*] *harritzen* *nau.*
 you here see.NMLZ.DET.ERG surprise.IPFV AUX
'Seeing you here surprises me.'

(175) Persian (Indo-European, Iranian: Iran; Mahootian 1997: 145)

[*Abejow-t-o* *xordæn-æm*] *to-ro* *æziyæt+kærd?*
 beer-2SG.POSS-OBJ drink.INF-1SG you-OBJ bother+did
'Did my drinking your beer bother you?'

---

[18] In regard to the verb forms, the difference in the proportion of dependent forms is significant in a Fisher exact test only if the two classes are compared directly, without any p-value adjustment ($p = 0.0364$ (> $p_{adjusted}$ 0.003)). This can also be seen if the whole domain of S-clauses is submitted to a Chi-squared analysis of the dependent verb forms, which turns out not to be significant (randomized $\chi^2 = 8.7669$, B = 100,000, $p = 0.1182$).

(176) Kwazá (isolate: Brazil, Colombia; van der Voort 2004: 698)

[*Mĩu* *'kui-a-nãi*] *nỹ-'dy-ja-ki.*
 chicha drink-1PL-NMLZ big-CAUS-INDEF.OBJ-DECL
'Drinking chicha makes fat.'

(177) Lavukaleve (Solomons East Papuan: Solomon Islands; Terrill 2003: 352)

Ta *hoina-ri-om* *a-na* *vo'vou roa* *a-na*
time.M DEM.SG.M-PSNV-M/N 3SG.M.OBJ-INCL boy.M one.SG.M 3SG.M.OBJ-INCL

hano mina uruala [*iru-i*] *ga* *a-o-kuru-ge.*
then *mina* properly sleep-NMLZ ART.SG.N 3SG.M.OBJ-3SG.SBJ-kill-ANT

'At that time, sleeping (i.e. sleepiness) was really killing one boy.'

(178) Tamashek (Afro-Asiatic, Berber: Mali; Heath 2005: 678)

I-kk`iræɖ=tæn [*a=d* *əjlə-n*].
3SG.M.SBJ-obligate.RES=3PL.M.OBJ DEM=COM go.SHRT:IPFV-3PL.M.SBJ
'It obligates them that they go.' (= 'They must go.')

In Dixon (2005: 164), A-clauses are generally defined as occurring with a CTP encoding a *primary* concept. As can be seen in the examples above, they always include a full lexical verb, with the causative component of meaning either lexicalized as part of the CTP or added as grammatical marking. Having said this, it is striking that the A-environment comes out as the least productive of all primary concepts in Table 20: It has by far the lowest CIP of all primary-concept environments, and a direct comparison with the next highest primary class (that of 'know-P') already shows a stark contrast (CIP_{lgs} = 69.1 versus 102.5); not surprisingly, this difference comes out a significant under a pairwise post-hoc comparison in the Kruskal-Wallis test. More generally, the attestation of A-clauses is cross-linguistically severely limited. In 59 of the sample languages, I have found no direct evidence for the occurrence of A-clauses in the sources available to me. In some of those, this is likely to be a mere gap in documentation. If we consider that, in languages whose grammar licenses the occurrence of A-clauses, these still tend to be rare in performance[19], it is very likely that this situation carries over to some of the other sample languages. Therefore, it cannot be ruled out that A-clauses are, in fact, constructible more widely.[20] However, this methodological obstacle notwithstanding, there is reason to believe that the restrictions on the productivity of A-clauses are empirically 'real': On the one hand, their rarity in performance even in languages where they are grammatically sanctioned is something that cannot be denied and calls for an explanation. On the other hand, we will see shortly there is clear-cut evidence from a number of languages that A-clauses are, indeed, an impossible construction type (or a 'constructional gap', as we said above). What we seem to be dealing with here, then, is an instance of Hawkins' (2004: 6)

---

[19] For example, Igor Nedjalkov (p.c.) reports that Participial complements in Evenki can be construed in A-function, but this is apparently rare in performance, so that it is hard to find pertinent examples in grammatical descriptions of Evenki. In his investigation of English Infinitives, Mair (1990: 23) finds that "the number of verbal [i.e. transitive, KSB] matrix predicates embedding infinitival subject clauses is not insubstantial […], although their frequency of occurrence is generally low."

[20] I refrain from providing an areal breakdown of the relevant languages here. Distributional issues will be taken up in §6.3.5 later on.

'Performance-Grammar-Correspondence Hypothesis', according to which there is a systematic correlation between the performance prominence of a given structure and its existence as a typologically productive construction type:

> When the grammar of one language is more restrictive and eliminates one or more structural options that are permitted by the grammar of another, the restriction will be in accordance with performance preferences: The preferred structure will be retained or 'fixed' as a grammatical convention, the dispreferred structures will be removed. Either they will be eliminated altogether from the output of the grammar or they may be retained in some marginal form as lexical exceptions or as limited construction types.

Taken together, the available evidence from performance and grammars suggests that the experiencer-object lexicalization pattern of primary two-argument verbs is significantly dispreferred in complementation to an experiencer-subject pattern. In other words, it is more likely for the complement proposition to be treated – quite literally – as an 'object' of thought, speech or experience rather than as the subject or the pragmatic starting point of the message. In my view, the guiding principle here is something that Siewierska (1993: 834) notes in connection with word-order preferences in the world's languages: "Human beings show an unqualified interest in themselves, their interlocutors and other humans. Consequently events and situations tend to be interpreted from the point of view of the persons involved rather than in terms of the events themselves or nonhuman or inanimate entities participating in these events" (cf. also DeLancey 1981, and more recently C. Everett 2008 for similar statements). Siewierska holds this principle responsible for the fact that a sentence like *People are dying of starvation* is generally preferred to *Starvation is killing people*, and for why *The horse was killed by lightning* is preferred to *Lightning killed the horse*. As can be seen, the preferred construals take the more animate participant as the linear starting point of the sentence, even though the inanimate stimulus is actually the 'antecedent' of the action expressed. The same effect on word order is also apparent where the stimulus is an S-clause or A-clause and the experiencer still surfaces clause-initially. We saw this at work in Chapter 2, with the corresponding example from Malayalam repeated here, and I add a similar one from Tariana:

(179) Malayalam (Dravidian: India; Jayaseelan 2004: 232)

En-ik'k'e [*Mary miDukki   aaNə   ennə*]   *toonn-i.*
1SG-DAT Mary clever.person be.PRS COMP   seem-PST
'It seemed to me that Mary is clever.'

(180) Tariana (Arawakan: Brazil; Aikhenvald 2003: 551)

Na-na   hiku-pana-pidana-sita-nha   [du-yami-hyu-pena].
3PL-OBJ   appear-ALL-REM:PST.REP-PRFV-PAUS   3SG.F-die-PURP.NVIS-NOMFUT
'For them it was obvious that she (her mother) was going to die.'

The diachronic relevance of such constellations was discussed in Chapter 2, where we noted that topicalized experiencers are commonly reanalysed as grammatical subjects (cf. Cole 1980 for the classic reference). In a similar way, Siewierska's principle also motivates the occurrence of subject raising in 'seem' and related predicates (*John seems to be ill*), and the personal construal of complement-taking predicates more generally

(recall our discussion of examples (165)–(167) above). But especially the latter indicate that we are not dealing with an issue of constituent order only. What is often affected by Siewierska's principle (and she does, of course, not fail to note this), is the mapping of referential properties and syntactic functions: As is well-known, syntactic functions tend to be sensitive to whether the referent in question is (in)animate, a speech-act participant, a specific or a generic entity, a definite or an indefinite one, etc., either in absolute terms or relative to the other participant in a two-participant scenario. The resulting grammatical effects have been captured by reference to an 'animacy', 'referential' or 'person' hierarchy (cf. Silverstein 1976 and many others; and cf. Bickel 2011: 404–412 for a recent synopsis of referential and scenario effects on syntactic functions). According to these mapping preferences, the S and especially the A function tend to be associated with the higher positions on the referential hierarchies and this, in turn, exerts a major limiting effect on the cross-linguistic productivity of A-clauses: A-clauses code propositions or states of affairs and are hence, by definition, inanimate discourse entities. Moreover, the non-subject participant in A-complementation is usually the opposite, i.e. an animate, often pronominal, entity, as can be seen, for instance, in (173)–(175) above. Given these absolute and relative referential properties of A-clauses, it is not surprising that their productivity in performance and grammars should be severely restricted.

This line of reasoning suggests that the way in which languages 'deal with' inanimate A-arguments in simple sentences influences the productivity of A-clauses. In order to examine this issue, I tried to obtain information, for each sample language, on inanimate A-arguments in simple sentences, particularly in hierarchical contexts (where an inanimate A acts on an animate P): Are these generally possible? Are they restricted to particular semantic types of inanimate A-arguments (e.g. natural forces)? And how are such restrictions related to the productivity of A-clauses in complementation? The relevant information is collected in prose format in a dedicated section of the database; the box is entitled 'Animacy constraints on subjects', and it can be found in the lower right area of every entry in the complementation data.[21] It is notoriously difficult to make this kind of (often very sparse) information available for precise cross-linguistic quantification. Some pertinent observations can, however, be made on a qualitative basis.

Overall, there appears to be a connection between the sanctioning of inanimate A-arguments and A-complement clauses, but this relationship is not always entirely straightforward. To be sure, there are languages which explicitly rule out inanimate A-arguments altogether, and those also normally lack the corresponding A-clauses. Thus we learn about Kiowa: "Whereas it is possible in English and other languages to use a transitive expression with an inanimate agent, e.g. 'the ice broke it' or 'the wind broke it', such constructions are not possible in Kiowa and must be rendered instead by coordinate clauses or by incorporation in the verb of the inanimate cause" (Watkins 1984: 111). Equivalent statements have also been made on Mapudungun (Fernando Zúñiga, p.c.) and Matsés (Fleck 2003: 831), where "there seems to be a restriction

---

[21] It goes without saying that retrieving this kind of information was not a trivial task, one that required substantial poring over reference grammars and other materials, or communication with experts. With regard to the latter, I am grateful again for the feedback and insights provided, mostly via email exchange, by a host of specialists.

against inanimate Ergative participants." This explains why not even the type of complement that is structurally closest to an NP, i.e. the nominalization, can function as an A-argument in Matsés:

(181) Matsés (Panoan: Brazil, Peru; Fleck 2006: 235)

    *[*Mua-ak-n*]   *ubi*   *nëish-me-e-k.*

     lie-NMLZ-ERG   1ABS get.mad-CAUS-NPST-IND

     'Lying makes me mad.'

Similarly, A-clauses in Kiowa and Mapudungun do not appear to be possible. For some other languages, we have similarly explicit information that A-clauses are not licensed in complementation. This holds, inter alia, for Yuchi ("A clause can only fill the patient role of a two-place verb." (Linn 2001: 496)), Ainu (Anna Bugaeva, p.c.)[22], Kolyma Yukaghir (Maslova 2003: 401ff.), Musqueam (cf. below), Semelai (Kruspe 2004: 348), Tariana ("A clausal argument cannot be in A function" (Aikhenvald 2006: 202)) and Yimas (cf. below). But while there seems to be a connection to limits to inanimate agents more generally in these languages, it is not completely predictable. For the Salish languages, including Musqueam, Kroeber (1999: 87) argues that "one generalization about the grammatical roles filled by subordinate clauses holds good throughout the [Salish] family. No subordinate clause functions as subject of a transitive predicate [...]. This fact is not too surprising, given that in Salish, inanimate entities of any kind are rare, though not quite impossible, as subjects of transitives." (Kroeber 1999: 87) In other words, here we do have evidence for the impossibility of A-clauses, but it does not fall out automatically from a general ban on inanimate A-arguments (cf. also Thompson 2012: 164, who argues against a direct connection in terms of animacy). A similar line of argumentation is offered by Foley (1991: 387) for Yimas: "Yimas complements can function in all core argument roles except A and [G]; these are forbidden because A and [G] participants must typically be higher animates and preferably humans, features complements typically lack." In Yimas, too, this is not a completely rigid constraint, though, because instruments can be grammatically construed as transitive agents ('the machete cut them' (ibid.: 203)); apparently, what is ruled out is more abstract inanimate As ('message, news'), and since these are semantically closest to A-clauses, this restriction may explain the ban of A-complements in the language.[23] In Ainu, Semelai, Tariana and Yuchi, there is evidence for the possibility of inanimate A-arguments, but it remains unclear how precisely this relates to the impossibility of A-clauses. Again, it may be the case that the usages licensed (in these languages, these are chiefly metaphorical agents like physical forces (e.g. wind, clouds, etc., or, as Kruspe (2004: 158) puts it for Semelai,

---

[22] Note that Bugaeva (2012: 498) writes that "complement clauses may have the syntactic functions of S, O, and rarely A", but upon closer inspection, it appears that those A-functions are limited to 'product' or 'headless relative-clause' readings of the nominalization in question (e.g. 'What you said surprised me'), while genuine complement interpretations ('That you didn't say anything surprised me') are ruled out. (This holds for the Chitose variety of Ainu; I have not been able to find comparable information on other dialects.)

[23] A similar but even more restrictive situation has been reported for Chimalapa Zoque: Johnson (2000: 114) states that the agents of transitive clauses can be "quasi-animate entities like the wind or the river", but that "instruments and sources are not allowed as subjects." This may explain the absence of A-clauses from Johnson's account of complementation, although there is no explicit statement that such constructions are ruled out.

"anthropomorphised inanimate entities, potent natural phenomena [...] and blades and bullets") are semantically still unlike the abstract mental causation underlying typical instances of A-complementation.

Supporting evidence for this latter position comes from Japanese and Vietnamese. In these languages, physical causation needs to obey the animacy hierarchy. Kuno (1973), for example, discusses the ungrammaticality of sentences like 'the accident killed the boy/him' in Japanese, and such sentences were also not readily accepted by my informant on Vietnamese. However, when the context changes to psychological causation, e.g. 'The news/message surprised/shocked me', all of my Japanese informants as well as my Vietnamese consultant showed no aversion at all. And it is telling that A-clauses are also acceptable in both languages.[24] Taken together, it may, therefore, turn out that the availability of *specific kinds of* inanimate A-arguments in simple sentences is correlated with the presence or absence of A-clauses, but an examination of this hypothesis requires a study in its own right. Let me just mention here that for most languages in my sample, the animacy hierarchy *can* be reverted in transitive clauses in specific contexts (unlike in Matsés or Kiowa).[25] Again, this assessment refers to the principal grammaticality of the relevant contexts, regardless of their frequency in actual performance; the latter is typically very low even where such contexts are grammatically licensed (for corpus evidence of more familiar languages, cf. e.g. Dahl and Fraurud 1996 on Swedish, or Bresnan et al. 2001 on English; exceptions will be discussed below). A common situation is thus like the following from Koyra Chiini: "Transitives which regularly have inanimate subjects operating on animate objects are rare, given the strong preference for animate subjects. However, when the inanimate referent is represented as an active agent", such as an illness or a force, "we can get such structures" (Heath 1999: 128), e.g. 'the disease afflicted you'. In sum, then, the availability of such constructions in simple sentences appears to be a necessary but not yet a sufficient criterion for the grammaticalization of A-clauses.

Given the constraints on A-arguments, and the amount of languages without A-clauses in the data, one may ask at this point what kinds of alternative strategies languages may employ for rendering the meaning associated with A-clauses. Since the primary meaning component of the A-environment is a causative one (change of mental or physical state induced by a state of affairs), a suitable alternative coding technique that comes to mind is that of a causal (or temporal) adverbial clause: 'It makes me sad that my hamster has died' = 'I am sad because/when my hamster has died'. This strategy is, in fact, reported for some of the languages in which A-clauses are known to be ungrammatical. In Yuchi, for instance, the translational equivalents of A-clauses are rendered by "adverb clauses", which "do not fulfill one of the required participants for the sentence but provide additional information" (Linn 2001: 501). This can be seen in the following example, where the subject of the matrix predicate is not the subordinate clause, but an animate participant (cf. the Agent indexation):

---

[24] In a mini survey conducted by my Japanese consultant Kyoko Maezono, all four of her fellow native-speaker informants accepted complements in *no* and *koto* in A-function (p.c.). For Vietnamese, a pertinent example was provided in (172) above.

[25] For an impression, the reader is again referred to the database, where the 'Animacy' box typically contains some relevant examples from each language.

(182) Yuchi (isolate: USA; Linn 2001: 502)

['*Ya-s'e-gwa-ci*]   '*œ-le*   *k'a-di-ta*   *jẽ.*
DIR-3SG(EM).AGT-sing-CLF.SUB:SIT   big-STAT thing-1SG.AGT-pride   PST
'It made me proud when/that he sang.'

A causal adverbial clause instead of A-clauses is also employed in Dolakha Newar
(Carol Genetti, p.c.), and a similar technique is clause chaining in Korafe (Farr 1999:
129). In general, the alternatives to A-clauses are hardly ever discussed in the available
materials, but it seems to me that adverbial clauses are *always* capable of conveying the
relevant meaning, so I suspect that these can step into the breach wherever A-clauses
are shunned for the reasons discussed above.

    After having established some major constraints on A-arguments and A-clauses in
the sample, I should also briefly make mention of languages that stick out by a
particular propensity for impersonal coding in the A-domain. In the literature, this has
been reported, by Evans (2004), for the Iwaidjan languages of Northern Australia. In
these languages, a "basic organizing principle" is to "project the chain of causation
directly such that causes or agents are represented by the subject, and effects or
patients are represented by the object, without being distracted by the ontological or
discourse properties of the cause or agent (human, topic, speech act participant, etc.).
[…]" (Evans 2004: 170). An example is provided in (183):

(183) Ilgar (Australian, Iwaidjan: Australia; Evans 2004: 170)

Ngan-ni-mi-ny   *wurrwiny.*
1P-3M.A-do-PST   shame
'I got ashamed.' (lit. 'Shame did me.')

Evans goes on to claim that "the dominant typological factor underlying the
development of experiencer-object constructions in the Iwaidjan languages seems to be
a high tolerance of active clause structures in which inanimate subjects act upon
human (and usually first person) subjects" (ibid.: 187), and that "experiencer-object
constructions are widespread in Australian languages, particularly in many head-
marking families of the north-west" (ibid.: 176). In my sample, comparable languages
are rather rare[26]; two candidates of this type are Amele and Yuracaré. Amele regularly
features impersonal constructions of the type 'it talls me', my head pains me', etc.
(Roberts 1987: 209). Their extension to A-clauses yields some interesting coding
effects, which is why these patterns will be postponed to §6.4 below. Therefore, let me
concentrate on Yuracaré here. In this language, a common pattern for coding human
experiencers in simple sentences is to treat them as direct or applicativized objects
rather than as subjects. Since experiencers are central in complementation, it comes as
no surprise that the coding preference for experiencer objects also pervades the
complementation domain in Yuracaré. As a result, a very prominent type of

---

[26] I am referring here to languages which are comparable to Ilgar in that the experiencer-object pattern is a
widespread and not an exceptional type of lexicalization. As was mentioned above, many languages do allow
occasional reversals of the animacy hierarchy. But as Malchukov (2008: 90) argues, "all experiencer object
constructions are subject to a functional pressure to syntactically upgrade the experiencer, which is the most
discourse-pragmatically prominent argument"; this pressure leads to the "diachronic instability of experiencer
object constructions" cross-linguistically (ibid.: 93).

complement clause is actually an A-clause. Depending on the syntactic type of object, we get an A-clause either in a canonical construction (with a direct-object experiencer, e.g. (184a)) or in a non-canonical one (with an applicativized experiencer, e.g. (184b); the latter is analysed by van Gijn (2006: 276) as instantiating Dixon's 'extended intransitive' clause pattern, which we subsumed under non-canonical A-frames in the present study (cf. §2.2 again and §6.4.2 below)):

(184) Yuracaré (isolate: Bolivia; van Gijn 2006: 319, 296)

    a. *Ti-jusu*   [*mi-n-nënë-ni=ti*].
        1SG.DO-want   2SG-APPL-cook-INT:1SG.SBJ=DS
        'I want to cook for you.' (lit. 'My cooking for you wants me.')[27]

    b. [*Abëssë-y=ti*]   *të-shujuta*   *ti-m-bë-ø*.
       play-1SG.SBJ=DS 1SG.APPL-sweat   1SG-APPL-treat-3
       'Playing football made me sweat.'

The patterns in Yuracaré are reminiscent of the S-prominent languages mentioned above, i.e. Jarawara and Wari'. In Wari', we also find A-clauses analogous to the ones from Yuracaré:

(185) Wari' (Chapacura-Wanham: Brazil; Everett and Kern 1997: 80)

    Hwaraman *na-pa'*   [*ca*   *com* *wa* *tamara'*].
    distasteful   3SG.SBJ.RL:PST/PRS-1SG.OBJ   CLIT sing INF song
    'The singing of a song is distasteful to me.'

In this example, the adjectival morpheme 'distasteful' receives full verbal coding by the so-called 'inflectional complex' following it; this indexes the Infinitival complement clause as the subject and the experiencer as the object of the main clause. In Jarawara, interestingly, the great inclination towards impersonal S-complements does not carry over to A-clauses: According to Dixon (2006b: 111), "the corpus includes very few instances of complement clauses in A function. All examples involve the causativized form […] of an intransitive verb." In other words, there is no system-wide preference for impersonal lexicalization patterns, and the language rather underlines the rarity of A-clauses in performance that we also encountered on a more global scale.

    The final issue to examine in relation to A-clauses is their position on the binding hierarchy, i.e. their structural preferences. As can be seen in Table 20, the A-environment is the third most desententialized group, outranked (in raw figures) only by the two phasal classes. With 57.9%, A-clauses have a relatively high proportion of dependent verb forms, and a correspondingly high mean CID (3.06); the boxplot in Table 20 shows that they join the CTP classes whose CID whiskers run the gamut from completely sentential to strongly nominalized constructions, and thus contrasts with classes at the upper end of the scale, whose whiskers never reach up that far into the desententialized domain. Under statistical inspection, again applying the non-parametric pairwise Kruskal-Wallis comparisons, it thus turns out that the median CID of A-clauses differs significantly from that of utterance, propositional-attitude and

---

[27] Note that my gloss APPL in the two examples subsumes different kinds of applicatives in Yuracaré, such as so-called 'Indirect-Object' applicatives, 'Cooperative-Object' applicatives, etc. They all promote an additional participant to a syntactic argument of the predicate.

knowledge complements, and exactly the same contrasts also hold for the proportion of dependent verb forms (with *p*-value adjustment). Less conservative estimates (e.g. direct comparisons without *p*-value adjustments) also bring out significant contrasts to the causative-P, jussive-P, DS-want-P and perception-P environments. In fact, qualitative evidence from a number of languages supports the apparent gravitation of A-clauses towards the deranking end of our binding scale. In these languages, there is even a crucial structural difference between S-clauses and A-clauses, to the effect that A-clauses must avail themselves of a considerably more desentialized type of complement than S-clauses. Good illustration of this constraint can be found in Hualde and Ortiz de Urbina's (2003) description of Basque. They state that "finite clauses in Basque […] seem to be ruled out from subject positions of transitive verbs. […] To express the intended meanings, nonfinite clauses may be used. […] There is no restriction for the subject sentences to the verbs or predicates that are not transitive" (ibid.: 648–50). This can be seen in the following examples:

(186) Basque (isolate: France, Spain; Hualde and Ortiz de Urbina 2003: 648, 650)

    a. *[*Faxistek boz gehiago atera dezaten(ek)*] *beldurra ematen dit.*
       fascists.ERG vote more win AUX.COMP.(ERG) fear give AUX
       'That the fascists get more votes frightens me.'[28]

    b. *Argi dago* [*Amaiak asko ikasten duena*].
       clear is Amaia.ERG much study.IPFV AUX.COMP
       'It's clear that Amaia studies a lot.'

The nominalized rendition of an A-clause was exemplified in (174) above. Basically the same constraint as in Basque can be found in Ndyuka and To'aba'ita: the more sentential complementation patterns can combine with evaluative-S and deontic-S predicates, but only the nominalization is capable of additionally expressing A-readings. In Fongbe, similarly, only the Infinitive can be used to render A-clauses, while S-clauses are attested in sentential constructions. In Ma'di, Lango and Supyire, the finite complements (Indicative and Subjunctive patterns) are extremely productive in the P-domain, but very restricted as subject clauses: They can act as an S-clause in a very limited fashion, and A-clauses cannot be covered at all in this way, but resort to the non-finite alternative, a strong nominalization or Infinitive.[29] Interestingly, though, Carlson (1994: 457) finds for Supyire that nominalized A-clauses are still infrequent overall, accounting for "only 13 examples in the entire corpus." This ties in with our earlier observation of the rarity of A-clauses in language use. Finally, a 'doubly

---

[28] The Ergative marker is bracketed here because it appears to be possible to omit it. The restriction against finite A-clauses is first formulated in the grammar as one against Ergative marking on such clauses, but it is added immediately that if the Ergative marker is removed, finite declarative clauses are still ungrammatical as A-clauses. Interestingly, the grammaticality "may improve, although marginally" (ibid.: 649) if the A-clause is an interrogative finite clause (e.g. 'How much money we got surprised us'), but their acceptability is still questionable.

[29] A nice comment on this distribution is provided by Noonan (1992: 192) for Lango: In this language, finite subject clauses "are unidiomatic except in the case of modals", i.e. some deontic-S expressions. But even those "impersonal constructions" are "notably rare; judgements are likely to be rendered personally." (ibid.) As an alternative technique, Lango speakers can resort to an Infinitive, and this is attested more widely in subject function; "Infinitives may occur in any sort of noun slot: They may be subjects, direct objects and objects of prepositions" (ibid.: 214).

relevant' example of structural constraints on A-clauses comes from Turkish. For one thing, a finite type of complementation can only be extended to S-clauses ('seem') but apparently not to A. And secondly, as was discussed in Chapter 4, Turkish has two productive nominalization constructions (*-DIK* versus *-mA*), which differ slightly in their degree of nominality (the latter showing some more nominal features than the former). The CTP distribution of the two constructions is very different and has been amply discussed in the literature, most recently by Csató 2010 and Johanson 2011, who claim that the functional difference largely boils down to the distinction between propositional versus predicational orientation. In keeping with this proposal, we find that *-DIK*-nominalizations are productive in the subject-clause domain chiefly for epistemic-S contexts and passive-S clauses derived from factive object-taking CTPs (e.g. 'it is known that …'), which are clearly propositional in nature. Evaluations are normally ruled out, but propositional 'be good' is an exception to that.[30] Crucially, A-clauses only occur in the *mA*-nominalization across all materials I have examined:

(187) Turkish (Altaic, Turkic: Turkey; Kornfilt 1997: 50)

[*Ahmed-in   sinema-ya   yalnız başına   git-me-si*]   *ben-i   çok   üz-dü.*
 Ahmed-GEN cinema-DAT alone   go-NMLZ-3SG 1SG-ACC very sadden-PST
'That Ahmet went to the movies by himself made me very sad.'

To the degree that *-mA*-nominalizations are more "noun-like" (Kornfilt 1997: 450) than the ones in *-DIK*, and hence also display a slightly higher CID in our coding scheme, Turkish is another language which supports a coding difference between S- and A-clauses.

In general, then, the observation is that several languages appear to make only the most strongly recategorized types of complement available to the A-function. Put another way, A-clauses tend to be particularly susceptible to analogical coding with NPs. Full-fledged clauses, and the propositions they code, provide the poorest match to the semantic and morphosyntactic prototype of transitive agents. As we saw earlier, these are normally highly individuated, animate, volitional instigators of controlled events (cf. Hopper and Thompson 1980 for these and other parameters). Moreover, they are especially prone to Givón's (1983) principle of 'topic continuity' (mentioned in Chapter 5) and the resulting economical coding; in discourse, most A-arguments are, in fact, pronominal rather than lexical and hence most likely not to be overtly coded by free forms at all (i.e. only by indexation) (cf. DuBois 1987). In terms of both semantics and morphosyntactic form, then, balanced A-clauses represent the exact opposite of what is prototypical and hence highly entrenched as an A-argument in any language. S-arguments, by contrast, are much more heterogeneous as far as the properties of their referents are concerned; among other things, they allow a much higher share of patientive and also propositional meaning, not unlike P-environments. This is, of course, reflected in patterns of morphosyntactic coding, such as 'semantic alignment', the best-known subtype of which is commonly called 'split intransitivity' (cf. Wichmann 2008 for a topical summary). And the same principle also extends to

---

[30] cf. Göksel and Kerslake (2005: 427): "Although *-DIK* clauses cannot, in general, be used to express evaluations […], the specific combination of *-DIK* with 'be good' does occur."

complementation. From the diachronic perspective developed in Chapter 5, an appropriate way of framing the issue is thus to ask why a finite/sentential complementation pattern, which typically arises in object-taking environments (e.g. quotatives, relative clauses, etc.), is more likely to be analogically extended to S- than to A-taking predicates. I believe that at least part of the answer lies in what has just been said, i.e. in the poor match of structurally sentential A-clauses to the many exemplars of A-arguments that a given language user has experienced in his linguistic biography.

In sum, the overall statistics collected in Table 20 suggest that A-clauses cluster most closely with predicational (rather than fully propositional) S-complements and those P-clauses that are preferably deranked in their structure (i.e. same-subject desiderative, emotive and phasal predicates). Individual languages may draw a structural line between S- and A-clauses, such that the latter can only be coded by overtly nominalized constructions, while the former also allow for complements whose structure does not involve recategorization.

6.3.5 Summary: General findings

In the last three sections I investigated the cross-linguistic productivity of different CTP classes in complementation as well as their statistical preferences for certain structural types of complement clause. As the sections had already been organized into P-, S- and A-environments, we can now draw some overall conclusions with regard to these three classes.

The empirical approach taken in the present study has revealed a principled correlation between different syntactic functions (S, A, P – broadly construed) and their likelihood of being coded by a complementation pattern. On the most general level, my data confirm Dixon's (2006a: 43) hypothesis that "complement clauses are likely always to function as O argument for some verbs. Depending on the language, they may also be in S and/or in A function." If the syntactic functions are understood in the broader sense introduced in Chapter 2, the statement can be made more precise in the following way: In 41 languages of my sample, both S- and A-clauses are attested. 44 of the sample languages are such that A is missing from the materials but there is evidence for S-clauses (with different degrees of productivity). And in 15 languages, neither S nor A is attested in complementation. Crucially, then, I do not have evidence for a language with A-clauses that does not also have S-clauses. Therefore, from a categorical point of view (i.e. disregarding the numerical productivity measure), there is a clear implicational pattern in the distribution of complements across syntactic functions: P>S>A. Put differently, if a language has A-clauses, it will also have S- and P-clauses; if it has S-clauses, it will also have P-clauses. This ranking can also be brought out statistically if we compare the CIPs of the three environments again: Across the data, the situation is always such that, for a given language, the highest CIP is found in one of the P-domains, followed by an S-environment, and finally by A (e.g. S=1, A=0.1, P=2.5). If those CIP 'peaks' are identified for S, A and P for each language, their means should again differ significantly. Fig. 14 below illustrates that this is, indeed, the case:
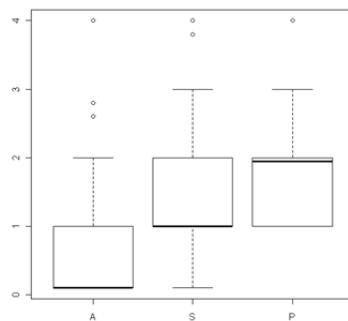
Figure 14. Distribution of the maximum CIPs for each function across the sample languages

A Friedman's ANOVA[31] reveals that the medians are significantly different overall (Friedman's χ^2 = 119.7, *df* = 2, *p* < 2.2e-16), and post-hoc tests show that the same is true for all pairwise comparisons (SA ≠ SP ≠ AP). In sum, the data suggest that S, A and P are differentially accessible to being coded by a complementation pattern (in much the same sense as different syntactic functions are also not equally accessible to being relativized, for example (cf. Keenan and Comrie 1977)). The major functional motivations for the P>S>A ranking have been discussed above, and there is no need to repeat them here. What I do wish to reiterate, though, is that the quantitative assessment needs to be taken with a grain of salt, especially the figures for A-clauses: Due to the rarity of such constructions in language use, they may not have found entrance into the grammatical descriptions of some of the sample languages, so the overall amount of languages licensing A-clauses may actually be more substantial than suggested by the figures just given. This also needs to be borne in mind when examining the areal distribution of languages without S-, A- and SA-clauses. A preliminary map is given in Fig. 15 below:
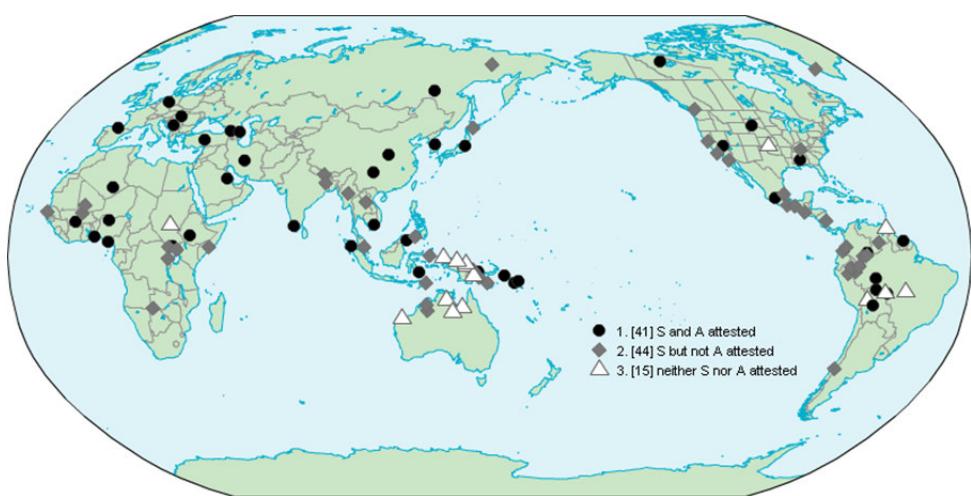
Figure 15. Areal distribution of sample languages with and without attested S- and A-clauses

---

[31] Friedman's ANOVA (cf. Field et al. 2012: 686ff.) follows exactly the same logic as the Kruskal-Wallis test we have applied so far: It is a non-parametric technique that compares the medians of more than two groups (i.e. an extended and non-parametric *t*-test). This time, however, we are directly comparing the CIP values for S, A and P for each language, so a test for dependent or paired samples is called for. Put differently, each language is treated like a participant that is measured under three different conditions, viz. S, A and P.

To the extent that the counts reflect the empirical reality, it appears that Australia-New Guinea and South America are the macro areas with the greatest lack of 'subject clauses' as a whole (cf. the white triangles).[32] The areal scope then expands considerably when A-clauses are taken into account, whose absence in the materials is quite common around the Pacific Rim, but also possibly extends to Southeast Asia and Africa (bearing our methodological problems in mind). If this distribution can be confirmed by future research, it would also mean that Eurasia (broadly construed) stands out as an area with higher average tolerance of A-clauses.

If we accept the above figures as valid for the moment, then another distributional question one may ask is the following: Given that languages tend to restrict the applicability of complement clauses in the hierarchical order P>S>A, and that a common coverage pattern seems to be {SP} to the exclusion of {A}, one may wonder whether the *inclusion* of the A-function is typical of those languages which commonly treat S and A as a grammatical category elsewhere in their grammar. Put the other way around, are languages with ergative categorization processes also those that draw an SP|A distinction in complementation? From a usage-based perspective, in which deeply entrenched categories form attractors or models for analogical extension, the underlying logic would be that complementation patterns are preferably extended from S- to A-function if this grouping is supported by other grammatical processes; conversely, if these other processes provide a model for the clustering SP|A, it may be that complementation patterns, which are fairly unsuited for the A-role to begin with, are preferably kept in this attractor basin. This hypothesis is not easy to test, considering that the groupings of S, A and P are always construction-specific and not a property of the language as a whole (cf. Kazenin 1994, Witzlack-Makarevich 2010, among many others). Languages may thus show ergative organization processes in various domains, and end up being 'ergative' or 'accusative' to varying degrees. But if we take some of the most overtly visible traits of this distinction, i.e. case marking of lexical NPs, verbal indexation and word order, a first approximation of this issue can be made. For each of the sample languages, we can determine whether the three grammatical domains of lexical case, indexation and constituent order show any vestige of ergativity. This can be done on the basis of the *WALS* chapters on the three parameters (Comrie 2011, Siewierska 2011 and Dryer 2011c, d), whose coding (among many other grammatical variables) was extracted and aligned for the larger project with which this dissertation is affiliated; the relevant information is collected in the database under the rubric 'Languages' > 'Grammatical profile'. For each of the three variables in question, one can translate the original *WALS* coding into a binary format that records whether or not the language exhibits signs of ergative organization (e.g. 'accusative' and 'neutral' case marking would be opposed to 'ergative' and 'split intransitive' (the latter of which aligns a subclass of S with P and hence is at least

---

[32] At the same time, it must be pointed out that the Australia-New Guinea macro area is known for its experiencer-object patterns in simple sentences (recall Evans 2004 from above for Northern Australia, and see similar remarks by Donohue (2008: 68) on New Guinea). In this light, future research on complementation in these languages may change the picture of A-clauses painted by Fig. 15 above, although it is also often noted that complementation is not well-developed in these languages to begin with (cf. §7.3).

partially ergative)).[33] One can then examine how languages with and without ergative categorization processes in these domains extend their complement clauses to S, A and P. This is set out in Table 21:

Table 21. Ergativity in lexical cases, indexation or word order, and the extension of complement clauses to 'subject' environments

	Both S and A	Only S	Neither S nor A
Ergative signals	6	20	7
No ergative signals	35	24	8
Total	41	44	15

The data in Table 21 are skewed quite heavily (randomized $\chi^2 = 10.61$, B = 100,000, $p = 0.005$), and a mosaic-plot inspection of the data shows that the strongest individual contribution to the skewing is made by the low proportion of 'ergative languages' which extend at least one of their complementation patterns to both S and A; apparently, then, this combination is significantly dispreferred (as, for example, in Matsés (cf. (181) above again)). The majority of languages with some ergativity in the above-mentioned domains extend their complements maximally to S (20/34 = 58.8%), while the majority of languages without ergative signals does, in fact, also include the A-function in complementation (35/67 =52.2%). Again, this does not mean that all complementation patterns of a given language can be used in this way, but that evidence for A-clauses has been found somewhere in the complementation system. Interestingly, though, it does *not* seem to be the case that the 'more ergative' a language turns out on the above parameters, the less likely it will be to develop A-clauses: It does not matter, for example, whether both case marking and indexation show ergative signals or whether only one of them does (Fisher exact test, $p = 1$). Clearly, however, much more extensive data are needed in order to investigate this issue properly. Considering that ergative alignment is a recessive type of organizing grammatical relations (cf. Bornkessel-Schlesewsky et al. 2008 for recent quantitative data from different perspectives), one would have to inspect a detailed sample especially of 'ergative languages' and their subject clauses. I leave this for future research to explore.

Another major conclusion from the previous sections can be drawn in regard to the morphosyntactic preferences of complements in S, A and P function. Although such preferences are best stated at the level of individual CTP classes, there also seemed to be some quantitative evidence to suggest that subject clauses, as a whole and on average, tend somewhat more strongly towards desententialization than object clauses. We saw that some languages make a distinction between S- and A-clauses, but there is further qualitative evidence from the sample that a significant line can also be drawn between S and P, or SA and P, respectively. A recurrent observation in unrelated languages is that only a subset of the complementation patterns that are used in object function can be used in subject function at all (S and/or A), and that this subset is

---

[33] So-called 'mixed' patterns typically also involve an ergative share, so they were included in the ergative coding here. In terms of word order, one may speak of ergative patterning when, in the basic order identified in *WALS*, S and P occur on the same side of the verb while A occurs on the other, e.g. AVP but VS, or PVA and SV (e.g. Mangarayi or Ungarinjin; cf. Dryer 2011d for discussion).

relatively more desententialized than the remaining constructions. This comes out most clearly where a complementation system shows a bifurcation between a highly sentential type of complement and one or more overtly 'non-finite' types, e.g. those with nominalized and related verb forms. The following languages are of this type, and in all of them, the sentential complement is blocked from the subject function, while the non-finite one can appear in at least S-function, and sometimes also A: Awa Pit, Barasano, Burmese, Hup, Jarawara, Korean, Lezgian, Mapudungun, Matsés, Motuna[34], Musqueam, Noon, Tukang Besi, Tümpisa Shoshone. (188) provides a relevant pair of examples from Korean:

(188) Korean (isolate: North Korea, South Korea; Sohn 1994: 56; Yang 1994: 241)

    a. *Minca-nun*   [*nwun-i*   *o-keyss-ta-ko*]   *sayngkakha-n-ta.*
        Minca-TOP/CNTR   snow-NOM come-will-DECL-COMP   think-IND-DECL
        'Minca thinks that it will snow.'

    b. [*Swunhi-ka*   *yeca-i-n+kes-i*]   *hwaksilha-ta.*
        Swunhi-NOM   female-be-NMLZ-NOM certain-DECL
        'It is certain that Swunhi is female.'

In other languages, there are several types of complement with non-nominalized verb forms, but other morphosyntactic properties (TAM, argument expression, nominal flagging) make one of them still less sentential than the other; and the resulting CID difference goes into the same direction as above: Only the more desententialized type has been found to function as a subject clause. Languages of this type include Choctaw, Lao, Menya, Santali, Tariana and Japanese. In the latter, complement clauses marked by *no* and *koto* historically derive from morphologically nominalized constructions whose dependent verbal morphology was reinterpreted as independent (cf. Wrona 2005, among many others), and a persistence effect betraying this nominal origin is the possibility for the complement subject to alter between Nominative and Genitive case, as shown in (189):

(189) Japanese (isolate: Japan; Wrona 2005: 123–24)

    Taroo *wa* [*kinoo*   *Ziroo ga/no*   *kita* *no*   *o*]   *siranakatta.*
    Taroo TOP   yesterday Ziroo NOM/GEN came COMP ACC   didn't know
    'Taroo didn't know that Ziroo came yesterday.'

Although the specific diachronic trajectories of *no-* and *koto*-complements are more different from each other than is commonly assumed (Wrona 2005, 2008), they share the synchronic property of possible *no/ga* case alternations and also their external flagging by appropriate case markers. This contrasts with complements in *to*, which show neither of the two properties and hence cannot be analysed as clausal nominalizations:

---

[34] In Motuna, a clause-chaining construction (illustrated in (152) above in a different context) may be used with mediopassive predicates that could be argued to take the switch-reference clause as their promoted S-argument (Onishi 1994: 416). This construction is excluded in the present study, and the contrast mentioned here is between the constructions for which Onishi explicitly argues that they function as complements of CTPs.

(190) Japanese (isolate: Japan; Kuno 1973: 213)

John   wa   [nihongo ga/*no   muzukasii   to]   itta.
John   TOP   Japanese NOM/*GEN   difficult   COMP   said
'John said that Japanese is difficult.'

Importantly, only the former constructions can be used as clausal subjects; the latter one, as it stands, is ruled out from this function. It can be made appropriate for it, though, if a more complex type of subordinator, again involving clausal nominalization (*to yuu koto*, lit. 'the thing that says that' (Kuno 1973: 218)), is chosen:

(191) Japanese (isolate: Japan; Kuno 1973: 213)

[*John   ga   sono   yuuwaku   o   kippari   sirizoketa   to+yuu+koto/*to*]
 John   NOM   the   temptation OBJ   resolutely   rejected   COMP

wa   migoto   da.
TOP   admirable   is

'It is admirable that John rejected the temptation resolutely.'

A very similar restriction is found in Malayalam, where the quotative construction can function directly as a subject clause only with the CTP 'seem'; in all other subject contexts, it needs to be augmented by a nominalizer:

(192) Malayalam (Dravidian, India; Asher and Kumari 1997: 46, 42)

a. [*Avan   satyasandhan   aaɲə   ennə*]   *ɲaan   viʃvasikkunnu.*
   3SG.M   truthful.person be-PRS QUOT   1SG   believe-PRS
   'I believe that he is honest.'

b. [*Avan   varum   ennatə*]   *aaʃcaryyakaram   aaɲə.*
   3SG.M   come-FUT   QUOT-**NMLZ**   surprising   be-PRS
   'It's surprising that he will come.'

The foregoing examples from Korean, Japanese and Malayalam have in common that the sentential type of complementation is historically derived from a quotative construction (Korean *ko* and Japanese *to* are quotative particles). In fact, this holds for almost all of the above-mentioned languages: a complementation pattern that is most likely to go back a quotative clause shows lexical diffusion maximally in the P-domain, but does not extend to the S- or A-domain, where alternative constructions have to be chosen. This restrictive pattern of analogical extension is striking, given how often it repeats itself in the data, but it is, of course, also possible for S-/A-taking predicates and utterance predicates to be united by the same complementation pattern. For example, in one language mentioned above, viz. Musqueam, they are both coded by the nominalization. Musqueam still obeys the principle under discussion, though, because there is a sentential type of complementation, historically derived from a conditional and/or imperative construction (Kroeber 1999: 248ff.), which cannot be extended to subject clauses.

Finally, in a third type of language, there are several constructions which involve dependent morphology of some sort, but again the ones that are more sentential by other criteria are not extended into the subject domain, while the ones with a higher CID are. This holds true of Wappo, for instance, where the relatively more finite (though still severely deranked) construction is productive in the P-domain, but only

the so-called Infinitive can be used in subject function. Examples of the two constructions were provided at several other places of the dissertation, but let me illustrate the relevant contrast again here:

(193) Wappo (Wappo-Yukian, USA; Thompson et al. 2006: 151, 140)

    a. *Ah*   [*te-ø*   *luče*   *poʔo-lah*]   *haṭis-khiʔ.*
        1SG.NOM   3SG-ACC   tobacco   smoke.DUR.DEP-NEG.DEP   know-STAT
        'I know that he doesn't smoke.'

    b. [*K'ešu*   *mehlah-ukh*]   *uwa-khiʔ.*
        deer   hunt-INF   bad-STAT
        'Hunting deer is bad.'

For strongly nominalizing languages like Quechua, subtle differences between subject and object clauses can emerge. Muysken (1994) reports that nominalizations in object function can alternate between Nominative and Genitive marking of the internal subject, while in subject function, only the Genitival marking is grammatical:

(194) Quechua (Quechuan: Peru; Muysken 1994: 2813)

    a. [*Xwan-ø/pa*   *hamu-sqa-n-ta*]   *yacha-ni.*
        Juan-NOM/GEN   come-NMLZ-3-ACC   know-1SG.SBJ
        'I know that Juan has come.'

    b. [*Xwan-*ø/pa*   *hamu-sqa-n- ø*]   *allin-mi.*
        Juan-*NOM/GEN come-NMLZ-3-NOM   good-DECL
        'That Juan has come is good.'

According to Muysken (ibid.), "the reason is presumably that in subject position, only noun phrases, but not clauses, can occur." Regardless of whether one wishes to refer to (194b) as a noun phrase or a complement clause, it is thus true that subject nominalizations show a slightly higher degree of desententialization than object nominalizations. The Quechuan case alternation is reminiscent of Japanese *no/koto-* complements, but the situation is somewhat different because in Japanese, it marks a watershed between two entirely different construction types (*no/koto-* versus sentential *to-*clauses), while in Quechua, it is a coding difference within the same constructional pattern, i.e. the nominalization. From this perspective, it would be interesting to find out whether Japanese *no/koto-*complements still show a statistical preference for Genitival subject coding when the complement is in subject function, while they tend to leave it in Nominative case when the complement functions as an object clause. This distribution would then replicate, as a performance preference, the conventionalized situation in Quechua. I have no Japanese corpus data on this issue, but the hypothesis would be intriguing to test in the future.

    What we have seen in several languages is thus a pattern of organizing complementation systems in such a way that only the overtly (or more clearly) desententialized members of the system can be used in subject function. In Chapter 5, we saw that the desententialization of subject clauses is often required by their position in the sentence: The strong inclination of subjects (in general) towards matrix-initial position, coupled with restrictions on the complexity of matrix-initial or left-branching material, imposes a pressure on many subject clauses to be either overtly decategorized

(e.g. being made non-finite or less complex) or recategorized (bearing NP marking of some sort, even if they are internally finite). However, the preference for desententialized subject clauses is not completely dependent on issues of constituent order but also holds beyond them. First, tight restrictions on the form of subject clauses can also be found in verb-initial languages, where both subjects and objects are postverbal. In Tukang Besi, for instance, complements can only act as subjects if they take the form of a strong nominalization; the sentential construction types are not possible here (Mark Donohue (p.c.); an example ((210)) will be discussed in §6.4 below). Second, the restriction on subject clauses also holds in languages where both subject and object clauses are preverbal and hence left-branching entities. Recall the above-mentioned nominal character of subject clauses in Japanese, which need not apply to all object clauses despite the fact that these are all preverbal, too.

Therefore, it appears that the pressure for desententialization is greater for subject clauses, regardless of their position. It was argued above that propositions show few similarities to the referents that languages normally select as subjects; this is especially true for A arguments but also certainly for S. Consequently, the predicates of a given language will be much more prone to take NPs rather than clauses as subjects. Frequency asymmetries of this kind are known to have certain cognitive effects, which in turn may help to explain the preferred morphosyntactic structure of subject clauses. Building on the usage-based literature, several such effects are theoretically conceivable. First, the marking of the most frequent and hence entrenched subject-predicate combinations is simply extended to instances in which the subject is propositional in nature. If subjects strongly tend to be NPs, for individual predicates as well as a generalized grammatical pattern, then other kinds of subject are easily drawn into this attractor state. This may explain, for example, why a VS/VO language like Tukang Besi simply replaces an NP subject by a strong nominalization after the same predicates: the succession of a predicate X and an ensuing NP is well entrenched, and so any other kind of subject following X is simply invested with the same coding. As Szmrecsanyi (2005: 113) puts it in a different context, "language users are creatures of habit", and 'habit' in this case does not relate to priming effects in the immediate context (as in Szmrecsanyi's study), but to long-term storage effects that influence the application of entrenched structural patterns to novel contexts, such as that of propositional subjects. Second, from the perspective of online planning and processing in communication, it has been argued that the unexpected constellation (here: subject = clause and not NP) needs to be marked overtly so as to signal the deviance to the hearer. In linguistic typology, this line of explanation has been invoked for differential object marking (e.g. Comrie 1989) and a host of other phenomena (cf. Croft 2003: Ch.6 for an overview). In relation to subject clauses, it may be another reason for the stronger overt nominalization of subject clauses as opposed to object clauses or, conversely, for why object clauses can shed some of the formal properties associated with NPs (such as the Genitival subject marker in Quechua). It also predicts that when a complementation pattern arises that is semantically and structurally very different from a canonical subject (such as a quotative construction), it will either resist

analogical extension to the subject domain[35], as in many of the above-mentioned languages, or it needs to be marked overtly to signal the deviation. This arguably applies to the nominalization of quotative constructions in Malayalam and Japanese we saw above. In this respect, the morphosyntactic differences between subject and object clauses tie in well with other coding asymmetries across the world's languages.

And finally, there is another respect in which these coding patterns are typical of cross-linguistic form-function hierarchies: They can have principled exceptions, which are normally also due to specific processes of diachronic change (this is also noted by Givón 1980 on his original binding scale). Thus in some of the sample languages, subject clauses are coded by a construction that is *more*, not less sentential than others in the complementation system: In Choctaw, Khwe, Koyra Chiini, Semelai, Tetun, Tzutujil and Urarina, the complementation pattern with the highest CID, which is often the only one with a clearly 'non-finite' verb form, is not used as a subject clause, but is found in the object-taking domain. In keeping with the binding scale presented in Table 20, one would expect that the P-environments affected should be the phasal and possibly emotive and same-subject desiderative classes. This is exactly what we find in the data. In all of the above-mentioned languages, the complementation pattern in question contains a converb, infinitive or bare stem (cf. §4.2 for these categories again), with equi-deletion processes applying to the internal subject, and the CTP classes affected are subject-control environments like phasal and same-subject 'want', or object-control contexts like 'help someone to do something'. A subordinator is either missing altogether or drawn from the semantic domain of goals (directional, purposive, etc.). (In other words, these are the precisely the constructions that undergo auxiliation in many other languages and thus effectively remove phasal and other same-subject environments from the domain of complementation on those languages.) By contrast, the complements of other CTPs in the above languages, including subject-taking classes, have developed along different historical lines: They might draw on the adverbial or more general switch-reference tracking domain (e.g. Choctaw, Urarina), on constructions that are also used as relative clauses (e.g. Koyra Chiini, Tzutujil) or simply on the insertion of a full-fledged unmarked clause into an argument position of the matrix predicate (e.g. Semelai, Tetun, Tzutujil). The synchronic outcome of these processes is illustrated by the following example from Tzutujil:

(195) Tzutujil (Mayan: Guatemala; Dayley 1985: 394, 392)

    a. *Xoqok*   [*chi tijoj*   *tiiʔiij*].
       1PL.ABS.began   to   eat.INF   meat
       'We began to eat meat.'

    b. *Najiini*   [*nakamsaaj*   *ja*   *tz'iʔ*].
       3ABS.is:in:progress   3ABS.3ERG.kill   the   dog
       'You are killing the dog.' (Lit. It's in progress that you kill the dog.')

It is patterns such as this in Tzutujil which contribute to the shape of the binding hierarchy as presented earlier.

---

[35] This is what Haspelmath (2008, to appear) calls 'differential selection' in diachronic change.

6.4 Syntactic functions of complementation patterns

In the preceding section, and especially in the summary just presented, I already dealt with the distribution of complementation patterns over the three major groups of syntactic functions, viz. S, A and P. However, it must be kept in mind that these were defined in Chapter 2 as rather broad comparative concepts which may gloss over the more specific syntactic categories found in individual languages. For example, our comparative concept of the 'P' function comprised direct and oblique objects alike. In Chapter 2, I thus introduced the distinction between 'canonical' and 'non-canonical' S/A/P arguments, and we have not yet incorporated such more specific syntactic functions into the picture. This is precisely the goal of the present section. It will be considerably briefer than the last one, as the issue will not be approached from a broad documentary perspective ('What kinds of syntactic functions are found in the individual languages?'), but from a more specific angle. As we shall see in due course, the primary question to be answered is, 'How widespread are canonical patterns in complementation?' This question is motivated slightly differently for subject and object clauses and thus needs to be operationalized and answered separately for the two environments.

6.4.1 Complementation patterns as objects

In the cross-linguistic literature, one can find different (though not necessarily contradictory) positions towards the object status of complement clauses. On the one hand, we have encountered Dixon's (2006a: 43) statement that "complement clauses are likely always to function as O argument for some verbs", where it is important to note that "O" in his terminology only refers to the direct object of a canonical transitive clause; other language-specific types of object (such as his 'E' or further ones that are not part of his classification to begin with) are not included here. On the other hand, we established in Chapter 2 that many of the typical CTPs (e.g. 'know', 'see', 'believe', 'want') are 'low-transitivity' verbs (in the sense of Hopper and Thompson 1980): they often involve experiencers rather than volitional agents, and the complement clause codes the corresponding stimulus rather than a clearly patientive, individuated object that is being acted upon and wholly affected by the verbal process. And even where they do involve a volitional agent and a complement that is not a stimulus, such as in utterance predicates, the situation is not a typically transitive one. In fact, it has been argued that utterance predicates, in particular, are often intransitive or at least not fully transitive across the world's languages (e.g. Munro 1982, Güldemann 2008). More generally, the literature on complementation is replete with warnings against treating especially 'finite' complement clauses as objects. In relation to English, for example, researchers have pointed out from very different angles (e.g. Deutscher 2000, Thompson 2002, Huddleston and Pullum 2002: 1017ff.) that *that*-complement clauses are often analysed as a single paradigm of 'direct object' clauses or 'noun clauses', based on examples such as (196):

(196) a. *I know* [*that you lied to your mother*].

     b. *I think* [*that you lied to your mother*].

     c. *I suppose* [*that you lied to your mother*].

All of these examples look basically the same, but this apparent transitive coding is somewhat deceptive: It is only with *know* in (196a) that the complement clause does appear in lieu of a direct object. *Think*, in contrast to *know*, does not take NP objects (**I think <u>it</u>*) but at best oblique objects (*I think <u>about it</u>*), although these create a significant meaning difference to the reading in (196b) above. Finally, *suppose* neither takes an NP nor an oblique object but does obligatorily take a complement clause or the pro-form *so* (*I suppose so.* but **I suppose*). For these (and other) reasons, all of the above authors argue that *that*-complement clauses in English should not uniformly be analysed as 'direct object' clauses. Instead, it must be acknowledged that they *can* fill the direct-object slot in the canonical transitive construction (as in (196a)), but that they may also function as P arguments (in our sense) in non-canonical environments.

Therefore, the primary question in the P-domain is to what extent we can find CTPs that are canonical transitive verbs, so that the corresponding complement clause can really be said to replace an "O" argument in Dixon's sense. In my sample, it turns out that nearly all languages have a least a subset of CTPs that are of this type.[36] It is only in three languages, i.e. Kiowa, Motuna and Ungarinjin, that I have not been able to identify clearly whether the complementation patterns in question can function as a direct object of a transitive verb. In Kiowa, there is only one construction that matches our comparative concept of complementation, and this construction is, as far as we can tell, right-adjoined to a full-fledged matrix clause (Watkins (1984: 235); cf. Dixon (2006a: 20) for the same verdict). The CTPs in question are either intransitive ('think') or transitive but with the addressee and not the complement coded as the direct object (e.g. 'tell'). In Motuna – again as far as I can tell from the description –, there are several complementation patterns, but they all have in common that, when used in P-function, they are not indexed as objects on the relevant CTPs, but instead fulfil a syntactic function that Onishi (1994) aptly refers to as 'complement': It is an argument position for these verbs, but one that is not treated canonically since it cannot be indexed like a proper object. This even holds for the nominalization construction, which is closest in form and function to a regular NP. Onishi (1994: 488f.) lists the syntactic functions of nominalizations, which include S-arguments and 'complements' but again not transitive objects (cf. (197a)). I have found examples in which such nominalizations 'complement' a transitive verb like 'see', but here it is the notional subject of the complement clause that is coded as the object of the CTP, with the nominalization being adjoined ((197b), cf. the discussion of 'raising' or rather 'control' in perception verbs in Chapter 4):

(197) Motuna (East Bougainville: PNG; Onishi 1994: 488, 416)

a. [*Tii   uri   pih-arei-ko*]   *haa-mu-ki-na.*
ART.LG home go-NMLZ-EMPH   want-**1S**-HAB.PST-F
'I wanted to go home.'

---

[36] Retrieving this kind of information involved analysing the argument-structural frames of individual predicates, scanning the descriptive materials for traces of transitive and intransitive coding, etc. In the database, the most important pieces of information can normally be found in the box "CTP distribution Comments". The actual variables (e.g. 'canonical-P', 'non-canonical-P', etc.), however, are not incorporated into the database in its present version. Their values were collected and submitted to analysis in a spreadsheet version of my data.

b. *... neeko-ng   koro=koro   mono-ong-juu   [ong   komikur-ei-ko].*
   1NSG.POSS.M language see-3P-1PC.A-CONT.DS DEM.M be.finished-NMLZ-EMPH
   '(It's worrying to) see our language dying out (lit. 'being finished').'

Note that it is possible to find the complementation patterns of Motuna as direct objects, but in all examples available to me, they have a 'product reading' (i.e. a headless relative clause like 'You can tell [what you are thinking of]') and not a processual reading as required by our definition of complementation. In Ungarinjin, finally, the situation is somewhat similar: one complementation pattern is restricted to a single intransitive verb root from which multiple complement meanings (quotation, causative, jussive, desiderative) may arise depending on the context; the other one, a generalized subordinate clause typical of Australian languages, has been found only as an S-clause and again in a product reading but not as a processual P-complement, but this could be a mere gap in the materials, after all.

For all other sample languages, then, it seems to be the case that at least one of the complementation patterns in the system can appear, in its genuine processual reading, as the direct object of a transitive verb, just as predicted by Dixon. In some languages, this transitive frame is extremely widespread in complementation. Thus in Semelai, virtually all CTPs, regardless of their position on Hopper and Thompson's transitivity scale, involve transitive encoding, as reflected by A-indexation on the CTP:

(198) Semelai (Austro-Asiatic, Mon-Khmer, Aslian: Malaysia; Kruspe 2004: 358)
       Ki=lən   [ki=ca].
       3A=want   3A=eat
       'He wants to eat (it).'

This is also explicitly confirmed by the description: "The complement functions in the S position in intransitive clauses, and in the O position of transitive clauses" (ibid.: 348). The same restriction of complement clauses to the canonical P-function is also found in Jarawara, Hmong Njua (cf. also Jarkey 2006 on an explicit statement on the closely related White Hmong) and at least certain varieties of Ainu (Bugaeva 2008a, 2012). And there are quite a few other languages for which canonical-P complementation is amply attested (but where intransitive CTPs or other non-canonical frames cannot be ruled out).[37] This even includes quotative environments, whose complement clauses can indeed be analysed as direct objects in several languages: Here, the situation is such that the utterance predicate could also appear with a direct-object NP as the *theme* (and not the addressee) argument. This is found, inter alia, in Fongbe, Jamsay, Malayalam, Martuthunira, Rama, Tariana, Tümpisa Shoshone, To'aba'ita, Tzutujil, Yuchi, Tamashek, Wambaya and Wolaytta. In the example from Tzutujil below, the addressee is coded as a prepositional phrase, while the complement clause is indexed as an Absolutive argument on the CTP:

---

[37] For many languages, for example, there is no coding difference between S and A arguments (which would be explicit evidence for an (in)transitive coding frame), and I could not determine whether all predicates that appear with P-complement clauses are also capable of occurring with a direct-object NP instead of the complement.

(199) Tzutujil (Mayan: Guatemala; Dayley 1985: 392)

Jar iixoq xb'ij chwe [chi ninb'e].
the woman 3ABS.3ERG.told to.me COMP 1ABS.go
'The woman told me to go.'

Where languages draw a coding distinction between S and A, the transitive patterning of utterance verbs can often be seen on the Ergative coding of the speaker (e.g. Lezgian, Kewa or Basque):

(200) Basque (isolate: France, Spain; Hualde and Ortiz de Urbina 2003: 452)

Jonek esan du [Mikelek erlojua galdu duela].
Jon.ERG say AUX Mikel.ERG watch lose AUX.COMP
'Jon said that Mikel lost the watch.'

But of course, the widespread attestation of at least a certain amount of transitivity in complementation must not be taken as evidence against the critical voices mentioned above. Quite on the contrary, because my data show equally well that the vast majority of languages also attest to scenarios in which a complementation pattern does *not* function as a direct object. Unequivocal evidence for such non-canonical P-clauses was found in 86 of the sample languages, which is still likely to be too conservative a figure. At the most coarse-grained level, there are two types of situation that lead to such non-canonical P-complements. In the first one, the CTP is syntactically intransitive, in the sense that it either never occurs with a non-subject NP in simple sentences, or with one that is not a direct object but an oblique argument. In Kolyma Yukaghir, for example, the desiderative predicate *erd'ī* takes intransitive indexation and, in this form, never occurs with an NP but obligatorily with a complement clause (Maslova 2003: 360). The complement thus never appears in lieu of a direct object, but itself acts as a non-canonical P-argument of the verb:

(201) Kolyma Yukaghir (Yukaghir: Russia; Maslova 2003: 415)

[*Čumu lejdī-din*] *erd'ie-j.*
all know-SUP want-INTR:3SG
'He wants to know everything.'

The list of languages with intransitive CTPs taking P-complements is long, including, for instance, Abun, Awa Pit, Barasano, Basque, Karo Batak, Burmese, Choctaw, Imonda, Kiowa, Krongo, Lao, Mapudungun, Matsés, Mosetén, Nkore-Kiga, Slave, Supyire, Taba, Tamashek, Tetun, To'aba'ita, Trumai, Tukang Besi, Tzutujil, Ungarinjin, Urarina, Wambaya, Yakan and Zoque. And as predicted by Munro (1982), a recurrent candidate for such intransitive predicates in these languages is a verb of speech. This includes situations in which the verb 'say' is commonly detransitivized when it occurs with a complement clause. This is the case, for example, in To'aba'ita and Yakan:

(202) Yakan (Austronesian, Western-Malayo Polynesian, Sama-Bajaw: Philippines; Brainard and Behrens 2002: 88)

Mag-aka iye [we' mag-kasuwa' ka'am].
INTR-say ABS.3SG COMP INTR-meet ABS.2PL
'He said that you will meet (each other).'

Non-transitive argument structures also include all cases in which either the subject NP or the object NP is coded in a non-canonical way, and where the complement clause appears in lieu of this object. Non-canonical subject marking typically signals that the subject has a less agentive role than usual, such as an experiencer; we saw these patterns in §2.2 for Lezgian (e.g. the Dative-Postelative case frame in (7)) and Matsés (the doubly Absolutive case frame in (6d)). If P-clauses occur with these predicates, they may thus be said to occur in a non-canonical construction:

(203) Lezgian (Nakh-Daghestanian: Azerbaijan, Russia; Haspelmath 1993: 365)

Wiri   ajal-r.i-z   [Musaq'.a-n   dide   q'e-nwa-j-di]   či-zwa-j.
all   child-PL-DAT   Musaq-GEN   mother   die-PRF-PTCP-SBST   know-IPFV-PST
'All the children knew that Musaq's mother had died.'

Conversely, non-canonical object marking means that the complement clause replaces an oblique rather than a direct object of the CTP. This pattern can surface as non-canonical case marking, such that the complement inherits the non-canonical case of the corresponding object NP.[38] This was illustrated in (64) (in Chapter 4) for Dative-complement clauses in Trumai, and a similar example is found with Ablative inflection on 'fear'-complements in Turkish:

(204) Turkish (Turkic: Turkey; Kornfilt 1997: 51)

Ben   [Ahmed-in   öl-me-sin]-den   kork-uyor-du-m.
1SG   Ahmet-GEN   die-ACT.NMLZ-3SG-ABL   fear-PROG-PST-1SG
'I was afraid that Ahmet would die.'

The pattern of oblique coding may also surface as adpositional flagging. Some CTPs do not take direct but adpositional objects in simple sentences, which are then replaced by complement clauses. This is, of course, most readily possible for nominalized complements, given their formal similarity to the phrasal NPs governed by adpositions. To name but two examples from the sample, the Infinitive in Lango and the nominalization construction in Lavukaleve are explicitly stated to be available as pre- and postpositional objects, respectively (Noonan 1992, Terrill 2003). Structurally sentential complements, by contrast, often cannot be governed by an adposition and may thus come to be attached directly to the CTP in question (creating the impression of a transitive pattern). In some languages, however, they may also be embedded as an adpositional object, as in the following example from Gulf Arabic:

(205) Gulf Arabic (Afro-Asiatic, Semitic: Kuwait etc.; Holes 1990: 21)

ʔasarr   ʕala   [ʔin   attasil   fiik   ʕala hal-mawđuuʕ].
insist.PRF.3SG.M   on   COMP   1SG.contact in.2SG   on   this-topic
'He insisted that I contact you on this topic.'

Apart from all these cases of non-transitive (including oblique-P) syntax, there is a second broad type of constellation for non-canonical P arguments. In these situations,

---

[38] Note that non-canonical case marking can, of course, also be found on complements of fully transitive CTPs, such as when a dative-marked purpose clause comes to fill the direct-object position of the transitive predicate 'want' and hence replaces an accusative-marked NP. But this is a slightly different perspective from the one pursued here: what we are interested in in the present section is instances of CTPs that are non-transitive to begin with, in the sense of requiring non-direct objects even in simple sentences.

the CTP *is* transitive, but the complement clause does not occupy the direct-object slot because this is already taken by an NP. This comprises (i) cases of utterance or other directive verbs that take the addressee, but not the theme as a direct object, with the complement being an additional argument; (ii) cases of monotransitive CTPs that involve raising or prolepsis of the complement subject into their own direct-object position; (iii) cases in which the complement is anticipated by an (optional) referential pro-form in the matrix, such as a demonstrative pronoun. Situation (i) is the mirror image of examples like (199) from Tzutujil above, which occurs, for example, in Mapudungun, Menya, Mosetén, Motuna, Rama, Turkish (CTP 'convince') or Noon:

(206) Noon (Niger-Congo, Atlantic, Northern Atlantic: Senegal; Soukka 2000: 274)

> *Ya   wo'-ee-ri   [an   jom-oo   daa   ki-maañ].*
> he   say-PST-OBJ.3SG   COMP   should-PRS.NEG   there INF-linger
> 'He told him that he shouldn't linger there.'

Situations (ii) and (iii) have been illustrated and discussed in different places of the dissertation (including (197b) above from Motuna), so we will not go into them again here. They have in common that the complement is adjoined to the matrix clause rather than being embedded as an argument of it.

In sum, a typical state of complementation systems is such that some CTPs in a given construction are canonically transitive, so that the complement clause replaces the direct-object argument. At the same, time, however, the complementation pattern is also applied to CTPs that do not take it as a direct object, either because they never select direct-object arguments to begin with or because some other NP fills this slot. Depending on whether or not the complement can then still be considered an argument of the CTP, these non-canonical patterns are either embedded or adjoined for the context in question. As was argued in Chapter 2, when the latter type of syntactic relationship, i.e. adjunction, holds for *all* CTPs that a given complement co-occurs with (rather than just a subset of them), we follow Dixon (2006a) in considering it a complementation strategy instead of a complement clause. In the P-domain, this holds for 13 constructions from 9 languages (Abun, Imonda, Kiowa, Martuthunira, Menya, Santali, Sanuma, Supyire, To'aba'ita). As can be seen, we have a rather strong bias to languages from the Australia-New Guinea macro area here, which will be returned to in Chapter 7. Note, however, that the majority of these languages also have a type of complement which can be embedded as an argument. For particularly engaging argumentations for the adjoined status of some complementation patterns, the reader is referred to Berry and Berry's (1999: 164ff.) discussion of *do*-complements in Abun and Lichtenberk's (2008: 984) discussion of so-called 'paratactic' complements in To'aba'ita.

6.4.2 Complementation patterns as subjects

For complementation patterns in subject function, the specific question we shall ask is a slightly different one. We already know from this and the previous chapter that subject clauses are structurally more constrained than object clauses: they are typically more overtly nominal in structure, to the point where only fully NP-like structures may be used in this function, and they may also not be allowed to occur in the

canonical subject position, being adjoined to the matrix and represented there by an overt or a covert pro-form. Given that both strong nominalization and syntactic adjunction are symptomatic of complementation strategies, what I wanted to find out is how widespread genuine complement clauses are in subject function. And in addition to structural and positional considerations, the above issue of canonical versus non-canonical coding also comes into play again, so one may ask even more specifically, How widespread are genuine complement clauses as subjects in canonical constructions? With this agenda in mind, we turn to S-clauses first before we examine A-clauses.

S-clauses were defined as complements filling the sole argument position of a monovalent predicate. Morphological coding constructions like case marking and indexation normally apply to a majority of S-arguments in the same way, and it is the behaviour of this majority class which will be referred to as the 'canonical' pattern. Where S-arguments are structurally heterogeneous in a less asymmetrical fashion, i.e. with two types of S as prominent patterns (such as in some so-called 'split intransitive' languages), we will assume two canonical coding patterns on a par. For example, if agentive-S arguments are systematically aligned with A while patientive-S arguments are systematically aligned with P, the P-pattern (which is the only relevant one for complementation) will not be said to be a non-canonical one. Non-canonical S-arguments rather arise where we find a certain subclass of predicates that exhibits some obvious differences from the way in which the majority of S-arguments work grammatically. For example, a common pattern is one where the CTP behaves differently from the majority of S-arguments with regard to indexation. Specifically, while the majority of monovalent predicates receive overt indexation, the predicates taking S-clauses do not. This is either because they are predicative adjectives without a copula verb or because they are verbal predicates that fail to trigger indexation. Such a lack of (otherwise dominant) indexation can be found, for instance, in Amele, Begak Ida'an, Barbareño Chumash, Korafe, Ma'di, Menya, Rama, Tetun, Yagua, Yakan and Yuchi. In the latter, there is normally overt indexation for both agentive and patientive S-arguments (according to a split-intransitive pattern), but no such indexation for adjectival predicates with S-clauses:

(207) Yuchi (isolate: USA; Linn 2001: 320)

[*Nehe shi-ci*]   *sẽ-:le*   *lœ*.
 here 2SG.AGT.arrive-SUB.CLF:SIT   good-VERY   ENC
'It's really good that you're here.'

What is also subsumed under non-indexed CTPs here is grammaticalized impersonal predicates that basically look like particles but do occur with an S-clause (e.g. impersonal 'must' in Koyra Chiini, Hungarian or Lao):

(208) Koyra Chiini (Nilo-Saharan, Songhay: Mali; Heath 1999: 295)

Bara [*ma   guna moŋgoro hun dooti*].
must 2SG.SBJ.SBJV see   mango   leave there
'You will undoubtedly see that mangoes are no longer there.' (lit. 'It must (be) that…')

In contrast to indexation, non-canonical case marking is much rarer (at least in my sample): Many clausal constructions (especially 'finite' ones) are barred from being

case-marked to begin with, and nominalizations normally take the canonical case morpheme (which is often zero-nominative or zero-absolutive in S-function). A possible candidate for non-canonical marking comes from Gooniyandi, where a Dative-marked purposive construction can co-occur with 'good':

(209) Gooniyandi (Australian, Bunuban: Australia; McGregor 1990: 400)

Mangaddi   jimandi   [ngab=goo].
not   good   eat=DAT
'It's not good to eat.'

It is unclear to what extent one can speak of a genuine 'subject clause' here (other relevant examples have not been found, which is also why the pattern was not considered in the sample in Chapter 5); on the assumption that *jimandi* is used as a monovalent predicative expression here, it seems best to treat the complement as a non-canonical S-clause.

From a global perspective, the situation that emerges from these considerations is such that 15 languages provide no attestation of S-clauses at all (as we saw earlier in this chapter), 77 languages show evidence of canonical S-complementation, and 8 languages of non-canonical patterns *only* (as far as I can tell from the materials available). These comprise the instance of non-canonical case marking in Gooniyandi, and unindexed CTPs in the remaining languages (Korafe, Menya, Tama, Tetun, Yagua, Yakan and Yuchi). One may, of course, question whether the unindexed nature of the CTP should really be taken as a signal of the non-canonicalness of the S-clause, and I have very little reservation about subsuming them under 'proper' S-clauses (cf. also Haspelmath (2011: 560f.) for the discussion of the typological treatment of the 'S' category). In fact, a much more severe problem than indexation lies elsewhere, namely in the structural and positional considerations referred to at the beginning of the section: The number of canonical-S languages decreases to 63 (and that of only non-canonical ones down to 6) if we subtract the cases in which the complements in question must all be considered complementation strategies: These are either too NP-like on structural grounds (and thus no complement clauses in the sense established in Chapter 2) or they must be removed from the subject position. For example, the S-clauses in Gulf Arabic, Ndyuka and To'aba'ita (to name but three completely independent languages) are exactly of this type: The sentential construction is obligatorily postverbal (in contrast to subjects in simple sentences), while the non-finite construction is basically a full-fledged NP and hence a very questionable instance of a 'complement clause'. In §5.5, I discussed the syntactic implications of obligatory removing the complement from the subject position. The loss of the subject status is clearest where the subject position is filled by a placeholder ('it', 'that'), which acts the grammatical subject of the CTP, with the complement being adjoined to a syntactically saturated matrix clause. Where no such placeholder is present, the situation is more debatable: We saw that it is commonly assumed that the complement is coreferential with a pro-dropped subject pronoun in the matrix; this approach leads to the same situation as above, i.e. an adjoined complement that does not act as the subject itself. On an alternative approach, one could argue that there is no covert subject pronoun, and that the complement has been removed from the subject position for independent reasons (e.g. weight), but still functions as the subject argument. The figure given

above (N_{lgs} = 63) assumes the more radical (but probably more accurate) position that *all* displaced S-clauses are deprived of their subjecthood.

In short, taking aspects of structure and linear order into account, one will have to conclude that a sizeable number of languages do not open their S-function to genuine complement clauses (in the Dixonian sense): While they have monovalent CTPs that co-occur with a complementation pattern, they severely restrict the form and the specific syntactic integration of these structures into the matrix clause.

The same remarks carry over to A-clauses, of course, as was suggested in this and the previous chapter. However, in this domain it is more worthwhile to first consider aspects of non-canonical coding, since the prototypical reference point (i.e. the canonical transitive construction) is more easily identified than in the domain of S-clauses and since there are quite a few interesting deviations from this prototype in the sample. A type of deviation that we have encountered before concerns the non-transitive marking of the experiencer object. This was illustrated with regard to case marking in German (cf. the difference between *beunruhigen* 'worry' with a canonical Accusative object and *gefallen* 'please' with a Dative object in example (10) in Chapter 2), and with regard to indexation in Yuracaré in the present chapter (cf. (184b) with an Applicative object index above). In some cases, indexation marks the verb as being fully transitive, but the syntactic status of the A-clause is debatable due to other reasons. In Tukang Besi, for example, my elicitation of A-clauses prompted Mark Donohue (p.c.) to provide the following example:

> (210) Tukang Besi (Austronesian, Western-Malayo Polynesian, Sulawesi: Indonesia)
>
> *No-mente-aku*   [*te*   *tinti-'a*   *nu*   *ana*].
> 3RL.SBJ-surprise-1SG.OBJ   CORE run-NMLZ   GEN   child
> 'The child's running surprised me.'

He comments that the nominalization functions as the "subject of a transitive verb" here (p.c.), and this would also seem to be confirmed by the facts of indexation on the CTP (i.e. third-person subject, first-person object). However, in a discussion of inanimate referents across several of the languages he has documented, Donohue (2004: 163) argues that in Tukang Besi, inanimate entities like trees (and other natural forces) are not able to occur as the A-argument of a transitive verb, in the sense that they cannot receive Nominative case (but the so-called Core case instead). This appears to also hold of the nominalization in (210) above, which is not marked by the Nominative but by *te* 'CORE' instead. Therefore, despite the fact that the verb structure fully conforms to the transitive construction in Tukang Besi, it seems that case marking does not, and so one could possibly speak of a 'non-canonical' A-clause here.

Amele has a similar pattern that removes the complement from the canonical prototype. To begin with, this language has canonical A-clauses, as in the following example provided by John Roberts (p.c.):

> (211) Amele (Trans-New Guinea, Madang: PNG)
>
> [*Hina ene*   *h-og-a*   *eu*]   *isiol-t-ena.*
> 2SG   here   come-2SG.SBJ-HOD   DEM   surprise-1SG.DO-3SG.SBJ.PRS
> 'That you came here surprises me.'

This complementation pattern ties in with the wide acceptability of inanimate entities as A-arguments in Amele (e.g. 'it talls me' for 'I am taller than it'). In fact, there is an additional construction type available for the expression of such scenarios, a so-called 'impersonal construction' discussed in great detail by Roberts (2001). There are several variants of this impersonal construction, but the important one in the present context is what Roberts calls the 'lexicalized' variant. It consists of three basic constituents: (i) a clause-initial experiencer NP occupying the canonical subject position; (ii) a clause-medial nominal like 'fear', 'desire', 'hunger', etc.; (iii) a clause-final impersonal verb that basically consists of an inflectional complex containing subject and object indices. Crucially, the experiencer NP is cross-referenced as the *direct object* of the construction, while the subject index always expresses an "anonymous third person singular" entity (ibid.: 226), similarly to a 'dummy' subject in other languages. An example is provided in (212):

> (212) Amele (Trans-New Guinea, Madang: PNG; Roberts 2001: 225)
>
> *Ija   wen   t-ei-a.*
> 1SG hunger 1SG.DO-3SG.SBJ-TOD.PST
> 'I was hungry (*lit.* I, x hungered me).'

The experiencer NP acts as controller of a host of syntactic operations (Roberts 2001: 245) and has thus many behavioural subject properties, but it remains a fact that it is construed as an object by indexation, and that there is an overt subject index in addition. Crucially now, such impersonal constructions can also combine with a complement clause, which yields the translational equivalent of desiderative and emotive complementation in other languages. Consider (213) below:

> (213) Amele (Trans-New Guinea, Madang: PNG; Roberts 1987: 47)
>
> [*Naus   uqa   na   ho   q-oc   nu*]   *ija   gale   t-ena.*
> Naus 3SG   of   pig hit-NMLZ for   1SG   desire 1SG.DO-3SG.SBJ.PRS
> 'I want (desire for) Naus to kill his pig.'
> (*lit.* I, x desires me, for Naus to Kill his pig.)

The critical question is, of course, for the syntactic status of the complement clause in such constructions: which grammatical function does it express? In Roberts (1987), such clauses are seen as so-called 'object complements', complementing the nominal expression ('hunger', 'desire') that is itself considered the object despite the anomalous coding on the verb. In Roberts (2001: 226), he shows that the stimulus need not be a nominal expression, but can also be adjectival or adverbial, and hence it cannot be an argument of the CTP. Therefore, "we need to posit an anonymous undefined subject" (ibid.: 229) as the index on the CTP. From this perspective, one may argue that the complement clause, while being syntactically adjoined to the matrix, actually spells out the stimulus of the desire, i.e. that it stands in apposition to the subject index on the impersonal verb complex, not unlike an extraposed A-clause of the English type *It bothers me* [*that she didn't take the dog for a walk*]. In fact, Roberts (2001: 246) maintains precisely that the subject index of the impersonal construction is a "place holder", which would make it similar to the 'dummy' subjects of extraposed subject

clauses we saw in Chapter 5. In this light, the impersonal construction in (213) above may well be seen as containing a non-canonical A-clause for comparative purposes.[39]

All cases mentioned so far thus testify to the well-known situation that stimuli and experiencers can differ from the canonical treatment of agents and patients in a two-argument clause. In Amele and Tukang Besi, this plays out in such a way that indexation is canonically transitive, but the stimulus is not fully canonical in either its coding (Tukang Besi) or its behavioural properties (Amele). In German and Yuracaré mentioned above, it is the non-canonical treatment of the experiencer that makes the pattern deviant from a regular transitive clause. More generally, where experiencers do not receive canonical case marking, it can be difficult to tell whether the stimulus should be seen as an object clause or a subject clause. In §2.2.1, we saw that what has often been called 'Dative subjects' in Malayalam is more properly analysed as oblique arguments in topical position, with the stimulus being the grammatical subject. The example provided earlier is repeated here for convenience because it crucially includes a complement clause as stimulus:

(214) Malayalam (Dravidian: India; Jayaseelan 2004: 232)

En-ik'k'e [*Mary miDukki   aaNə   ennə*]   *toonn-i.*
1SG-DAT Mary clever.person be.PRS COMP   seem-PST
'It seemed to me that Mary is clever.'

A similar analysis may apply to the following structure from Tariana. For this language, Aikhenvald (2006: 189) argues that object-coded experiencers are non-canonically marked subjects and that if the stimulus is a complement clause, it "fills the O slot":

---

[39] A very similar construction is also found in Motuna, with the verb *haa* 'want'. Interestingly, Onishi provides different analyses for this construction across his publications. In Onishi (1994, 2000), we find this predicate analysed as an 'extended intransitive' one (in Dixon's sense): It indexes the experiencer as S-argument (in patientive coding according to a split-intransitive fashion) and leaves the stimulus as an unindexed 'complement'. This is shown below and corresponds to the analysis of complement clauses given in (197a) earlier on:

Motuna (East Bougainville: PNG; Onishi 2000: 134)

Nii   tuu   haa-mu-u-ung.
1SG   water   want-S-IRR.PST-M
'I want water.'

In a later publication, however, Onishi (2004: 85) states that the predicate *haa* is actually an "impersonal transitive verb" which is indexed for both an Undergoer and "the invariable third-person A suffix *-u*". The same example as above is thus now glossed in the following way:

Motuna (East Bougainville: PNG; Onishi 2004: 85)

Nii   tuu   haa-m-u-u-ung.
1SG   water   want-1O-3A-IRR.PST-M
'I want water.'

As can be seen, the critical difference lies in the analysis of the marker *-mu*: It is seen as a monomorphemic S-index in the first analysis, but as consisting of two morphemes, *-m* and *-u*, in the second analysis. Crucially, however, the status of the NP 'water', or of the complement clause in (197a) above, would also change with the analysis adopted: It would be a non-canonical P-argument on the first interpretation, but a non-canonical A-argument on the second one, exactly parallel to what is going in Amele. For the quantitative picture to be presented in due course, a non-canonical A pattern was thus recognized for both Amele and Motuna.

(215) Tariana (Arawakan: Brazil; Aikhenvald 2006: 189)

[*Du-haniri-ne*   *pa-sape-hyu*]   *manhina-ma-na*   *nu-na.*
 3SG.F-father-INS   IMPS-speak-PURP.NVIS   be.hard-EXCES-REM:PST.VIS   1SG-OBJ
'It was hard for me to speak with her father.'

As Aikhenvald (2006: 181) herself concedes, "the subjecthood of the [experiencer] constituent is somewhat problematic", as it partakes in far fewer subject-sensitive operations than, say, Dative experiencers in Lezgian (Haspelmath 1993). The latter are subjects in so many respects that an analysis of the complement as (non-canonical) A-argument seems contrived; for Tariana, by contrast, it is at least a possibility.[40]

Finally, example (214) above from Malayalam shows another type of complication in regard to non-canonical A-clauses: For some predicates (here: 'seem'), it can be hard to decide whether the experiencer is an argument to begin with or rather an adjunct; put differently, is the CTP monovalent or bivalent? For 'seem' in Malayalam, Jayaseelan (2004: 241) puts forward a bivalent analysis, which makes it possible to speak of an A-clause in (214) in the first place; on a monovalent analysis, the complement is in S-function, and the experiencer an adjunct NP. Similar problems arise in Slave, where one of the mirative predicates ('surprise') takes a complement clause as the stimulus and has the experiencer coded as an oblique NP:

(216) Slave (Na-Dene, Athapaskan: Canada; Rice 1989: 1249)

[*Mary*   *kee*   *wehsị*   *nị*]   *segha*   *yedodíya.*
 Mary   slippers   3.made   COMP   1SG.for   3.surprised
'It surprised me that Mary made slippers.'

While it is uncontroversial that the CTP is intransitive in the sense that it cannot take a direct object, it is less clear whether the experiencer has argument or adjunct status. Slave has a rather productive class of what Rice calls 'oblique objects', and if this carries over to the experiencer in (213), the complement would be a (non-canonical) A-clause respectively. And the same holds for Tariana again, where Aikhenvald (2003: 551) argues for the complement clause of 'seem/appear' to be an intransitive subject (i.e. 'S' in the typological, not the monovalent, sense):

---

[40] Another such case is found in Korean, where predicates of 'fear' appear either in a doubly Nominative frame, or with the stimulus in the Nominative and the experiencer in the Dative (cf. Yang 1994, Gamerschlag 2007). The facts from case marking thus suggest an analysis of the stimulus as being at least equal on the hierarchy of syntactic functions. Accordingly, when the stimulus is coded by a complement clause, one may thus be able to speak of a non-canonical A-clause:

Korean (isolate: North Korea, South Korea; Gamerschlag 2007: 90)

Chelswu-nuni   [*Mina-ka*   *Yenghi-lul*   *tasi*   *manna-nun*   *kes-i*]   *twulyep-ta.*
Chelswu-TOP   Mina-NOM   Yenghi-NOM   again   meet-PRS.REL   NMLZ-NOM   fear-DECL
'Chelswu fears that Mina meets Yenghi again.'

Note that the centre-embedded position of the complement clause is less criterial than in Lezgian again, as A-arguments may also show up in this position and are less tightly constrained to the sentence-initial position than in Lezgian. Therefore, the specific constellation in the example above may also fall under a comparative concept of non-canonical A-clauses, although I admit that this is an adventurous proposal. For analysing the complementation system in Korean, nothing hinges on it, though, as the above complementation pattern can also be used as a clear-cut (i.e. canonical) A-clause (cf. Park 1995). (Because of its highly debatable character, the structure exemplified above was also *not* included in any of the quantifications on A-clauses, including the one to follow in the present section.)

(217) Tariana (Arawakan: Brazil; Aikhenvald 2003: 551)

Na-na   *hiku-pana-pidana-sita-nha*   [*du-yami-hyu-pena*].
3PL-OBJ   appear-ALL-REM:PST.REP-PRFV-PAUS   3SG.F-die-PURP.NVIS-NOMFUT
'For them it was obvious that she (her mother) was going to die.'

If these examples are included for now under two-argument constellations and hence count as non-canonical A-clauses, the overall situation in my data is as follows: In keeping with earlier calculations, there are 59 languages without any type of A-clause being attested; there are 36 with canonical A-clauses, and 5 in which *only* non-canonical ones have been found (Begak Ida'an, Motuna, Slave, Tariana, Tukang Besi). Languages with both canonical and non-canonical A-clauses include Amele, Georgian, German, Malayalam, Yuracaré and possibly Karo Batak[41]. Given that our primary interest is with canonical A-clauses, it is again important to note that their overall number decreases if complementation strategies are removed from the sample. If strong nominalizations and extraposed/adjoined complements are discarded (e.g. Gulf Arabic, Hmong Njua, Jamsay, Ma'di, Ndyuka, Supyire, To'aba'ita), there remains a total of 29 languages in which it appears to be possible to open the canonical A function to genuine complement clauses. Again, I must emphasize that there may be vast gaps in the documentation of A-structures in the sample languages, but the present section has shown again that, at any rate, the canonical A-function is considerably more constrained than P and S in complementation. This confirms, on a syntactically more specific level, the observations made in the present and the last chapter, and thereby concludes our analysis of the grammatical properties of individual complementation patterns. In the final chapter of the dissertation, we will look at the historical trajectories and synchronic organization of these patterns in complementation systems.

---

[41] In Karo Batak and Begak Ida'an, there are two equally important voice systems (Actor/Undergoer voice) and complements are often "grammatical subjects" (cf. Gooudsward 2005: 338). Depending on the specific analysis adopted, some of them may best be analysed as A-clauses, given that the Undergoer voice does not demote the Actor to adjunct status but leaves it as a core argument (cf. also fn. 16 above). I have counted the two cases as non-canonical A-arguments.

7

Emergence, diffusion and systemic organization of complementation patterns

7.1 Introduction

The final analytical chapter of the dissertation continues to examine the co-occurrence patterns of complements with different CTP classes, but with a shift in perspective. This time, our attention will be directed mostly at the complementation patterns themselves and the functional profile they establish by distributing over the complement-taking environments in a specific way. For example, one type of complement in Wambaya co-occurs with same-subject desiderative, jussive, causative and emotive predicates like 'happy/sad', but is ungrammatical with phasal, perception, knowledge and utterance predicates, for which other types of complement are reserved. As was argued earlier, such distributional facts are the product of lexical diffusion processes, driven by analogical extension over time. In §7.2, I would like to take this perspective one step further and ask whether complementation patterns with similar CTP profiles in my data are also likely to derive from similar source constructions in the first place. The underlying assumption is that, once a complement arises in a particular context, it is analogically extended in a pattern that one could liken to Wittgenstein's (1953) family resemblance model: its diffusion is guided by syntactic and semantic similarities to the original context or 'exemplar', and can become constrained if no sufficient similarity is recognized by the language users. In other words, it proceeds from the least conspicuous contexts to more 'salient' ones, using the terms of de Smet's (2012) recent comprehensive paper on this issue (cf. also Timberlake 1977 for similar constraints on analogical extension).[1] As a result, one may assume that the synchronic CTP distribution is still to some extent a reflection of the

---

[1] The family-resemblance model is, of course, also the central idea of the well-known semantic map approach in linguistic typology (cf. Croft 2003: Ch.4), and the specific CTP profile of a given complement is a perfect example of a semantic map.

original bridging context, a persistence effect in Hopper's (1991) sense. This is what I intend to explore in more detail; in the course of §7.2, we will thus probe into the diachrony of clausal complements and attempt to relate historical sources to synchronic CTP distributions.[2] In a much briefer section §7.3, the chapter will then be rounded off by exploring typological differences in how complementation patterns are organized into complementation systems.

7.2 Emergence and diffusion

As was stated above, the specific functional profile of a given complementation pattern can be seen as a synchronic result of gradual and principled diffusion from a historical source. In a comparative study, these processes can be looked at from two perspectives. On the one hand, we can investigate the cross-linguistic clustering of complements with a similar CTP profile, and ask for their diachronic or functional 'background'. This is the perspective taken in §7.2.1. On the other hand, the synchronic distributions can tell us something about the CTP classes themselves, viz. how they are selected – or "voted for" (Croft and Poole 2008: 10) – in the diffusion process of complementation patterns. Therefore, the same data can be used to create a bottom-up classification of the CTP environments of the present study, and ask for the functional principles by which they cluster together. This is the perspective taken in §7.2.2.

7.2.1 The cross-linguistic clustering of complementation patterns

There is a vast body of literature on the emergence and change of grammatical constructions from a functional and usage-based perspective; prominent monographic treatments with a cross-linguistic orientation include Harris and Campbell 1995, Croft 2000, Hopper and Traugott 2003, Heine and Kuteva 2007, Fischer 2007 and Bybee 2010, to name but some of the most widely cited works. In equal measure, there is an extensive literature on the diachronic dynamics of complementation patterns, comprising both philological and cross-linguistic approaches (pertinent references will be provided throughout this section). In analysing the data for this study, I have attempted to bring this body of research to bear on the characterization of the individual data points of my sample. For each complementation pattern, I thus tried to identify traces of historical origins and developments as far as possible. For the better-known languages with a relatively long written tradition and lively philological research, I was fortunate to be able to draw on published historical research and on the availability of etymological dictionaries. The latter were exploited as a source of information on the subordinating morpheme, which remains one of the best indicators of the history of a given complementation pattern. I am aware of the problems involved in relying on such subordinators for reconstructing a possible historical origin

---

[2] This idea is also, albeit more indirectly, foreshadowed in Dixon's (2006a) overview of complementation, where he proposes some "tentative associations" (ibid.: 43) between certain types of his "complementation strategies" (such as clause juxtaposition, relative clauses, purpose clauses) and typical complement-taking predicates. To the extent that his complementation strategies are often simply the diachronic source constructions of complementation patterns more generally, the present chapter is thus fleshes out and extends Dixon's suggestions.

of a whole construction (this is called the 'Marker/Structure Fallacy' by Harris and Campbell (1995: 284)). In the database, one will thus find two separate fields of information on diachronic issues, one relating to the subordinator as such (e.g. its synchronic multifunctionality and, where available, information on its etymology), and a second one on possible diachronic scenarios for the construction as a whole. For most of the sample languages, direct diachronic evidence is missing altogether, and so it has to be conceded that I am skating on ice that is just as thin (or thick, for that matter) as in other cross-linguistic studies of grammaticalization processes (cf. e.g. Heine and Kuteva's (2007) adventurous endeavour to reconstruct language evolution by recourse to grammaticalization). In these cases, then, what we can examine directly is patterns of synchronic constructional polysemy, and these can in turn be taken as being suggestive of certain diachronic developments that are known to have been recurrently operative in languages with actual historical records. Despite this inherent methodological problem, it should also be pointed out that there is a lot of consensus in the literature on the most common sources of complementizers and complementation patterns, and at least some of the relevant diachronic hypotheses are not very controversial; for example, where a complementizer is transparently related to a verb of speech, there is little doubt that the construction emerged via a so-called quotative pathway (e.g. Deutscher 2011) and did not, for instance, emerge by the reanalysis of a relative clause. In sum, the overall idea of the present section is to see the CTP distribution of complement clauses in the light of constructional polysemy and the potentially underlying diachronic processes.[3]

We shall begin, of course, by establishing which complementation patterns in the data exhibit a similar functional profile. This can be tackled empirically by turning to (dis)similarity analyses again. But in contrast to the ones we performed in Chapters 4 and 6, where the distance between languages or CTP classes was mapped out based on a few numerical variables, the present analysis takes a different conception of statistical similarity as its starting point (cf. Gries 2013: 343ff. for discussion): Specifically, we want to compare how each of the complementation patterns 'behaves' in regard to the various CTP classes, i.e. what specific kind of co-occurrence relationship it contracts with them. Following the coding presented in Table 18 of the last chapter, a given complementation pattern may, for example, achieve full coverage (= '1') with utterance and propositional-attitude predicates, be grammatical but dispreferred with knowledge predicates (= '0.9') and be ungrammatical with phasal and desiderative CTPs (= '0'); this example is fictitious and is merely meant to illustrate the point. As a result, each complementation pattern can be thought of as a directed vector with as many dimensions as there are CTP classes, and it basically takes the shape of a string of numerical values {0; 0.5; 1; 1; 1; 0.9; 0; 0; …}.[4] What we are now interested in is which vectors exhibit a similar overall shape (or 'curvature' in statistical terms) because their

---

[3] For another recent study of constructional polysemy in complex-sentence systems, the reader is referred to van Lier (2009). This work explores a very different aspect of the multifunctionality of dependent clauses, viz. its dependency on the flexibility of word-class boundaries in a given language.

[4] For the calculations, I considered all predicate classes distinguished in Chapter 6, except for 'other-S' and 'other-P' since these are too heterogeneous in nature to make any meaningful contribution.

high and low grammaticality scores distribute similarly across the CTP classes in the sample. Therefore, an appropriate similarity (rather than a distance) measure has to be chosen. A popular measure of vector similarity in psycholinguistics is the so-called 'cosine' (cf. Manning and Schütze 1999: 299ff.), and this is the kind of statistic I applied in order to create a similarity matrix of our data.[5] This metric can then be submitted to the *SplitsTree4* software introduced in Chapter 4, which performs a specific cluster analysis of the data at hand and visualizes the results as a spatial representation. The 'raw' version of this so-called *NeighborNet* arrangement of the data can be found as Material 7 in the Appendix, while Material 8 labels some of the distinctive clusters that will be relevant in the following discussion.

As can be seen in Material 7, what we get is a star-shaped display of the data, with individual data points reaching out into all kinds of direction and with many individual lines. This reflects the fact that each complementation pattern defines a unique CTP profile, which is only expected if we consider that it is part of a highly specific grammatical ecology: As in any complex adaptive system[6], its profile is determined by the existence and nature of other complements in the language, or by alternative means of expression that the language may resort to (cf. Chapter 6 again, and also §7.3 below). For this reason, it is comprehensible that hardly any pair of complementation patterns in the data shows exactly the same CTP distribution. However, there is also a fair amount of 'networking', or reticulation (Bryant and Moulton 2004), going on (average delta score = 0.405, *Q*-residual score = 0.11), and Material 8 highlights some of the larger clusters that can be detected (though no sharp boundaries can often be drawn); for ease of reference, these have been labelled clockwise from A to I. Those clusters, then, represent groups of complementation patterns that exhibit a fairly similar distributional profile, and we are now going to explore whether they also tend to contract similar patterns of constructional polysemy, or share certain features that could be indicative of their diachronic origin.

To start with, let us turn to cluster C in the upper right corner of the network. The complementation patterns in this area have a CTP profile that primarily comprises what Givón (1980) calls 'cognition-utterance' verbs as P-clause-taking predicates. Constructions that are found towards the root of this cluster typically cover utterance, propositional-attitude, emotive, knowledge and sometimes also direct-perception predicates (e.g. Awa Pit1, Kewa, Fongbe1, Abun1)[7]; but the further one moves outwards along the branches, the more the CTP coverage becomes restricted to utterance predicates (e.g. Tümpisa Shoshone2, Mapudungun3, Martuthunira4, Matsés4, Krongo2), and, crucially, there is no (or extremely little) extension into the

---

[5] The computation draws on the routine `Dist` implemented in the package `ape` for *R*.

[6] On the notion of 'complex adaptive systems' and 'dynamic systems theory' in the behavioural sciences or linguistics more specifically, cf., for example, Beckner et al. 2009, Cooper 1999, Dörnyei 2009 (Ch.3), Holland 2006, Howe and Lewis 2005, Tabor and Tanenhaus 1999, van Geert 2008 and Wildgen 1984.

[7] Throughout this chapter, the numbers given behind individual complementation patterns are part of the "ID" of that construction in my database (e.g. "Awa1"), and it is these abbreviations that are also found in the *NeighborNet* display of the data. Therefore, the numbers are intended to assist the reader in identifying which data points in the network I am talking about.

domain of subject clauses (e.g. evaluative or deontic predicates). Given this particular CTP profile, one would expect that the constructions in question probably emerged in a **quotative** context, and this is borne out very powerfully: Most of the patterns in this area can be reliably traced to a quotative origin and, conversely, the constructions which have a quotative source in my data are indeed predominantly found in this cluster and only exceptionally in other areas of the network (we will return to those below).

The emergence of complement clauses from quotative constructions is one of the best-studied instances of grammaticalization in the literature (e.g. Ebert 1991, Lord 1993, Klamer 2000, Deutscher 2000, 2011, Güldemann 2008, among numerous others). The typical scenario for this type of diachronic change involves the concatenation of three elements: a quote, "a clause that means roughly '(X) says/said', and the regular appearance of this clause in paratactic sequences after verbs of speech. From there, the road to reduction is open." (Deutscher 2000: 90) In other words, the original sequence is one such as 'I spoke, I <u>said</u>, "[quote]"'. It is the underlined part that becomes reanalysed as a quotative marker and finally a more general complementizer. I am, of course, vastly oversimplifying here, but since the pathway in all its details has been so well-described, this is probably sufficient. However, let me return to two important points in Deutscher's statement above: One is his formulation "a clause that means *roughly* 'X says/said'". As Deutscher and especially Güldemann (2008) show, the linking element that ultimately grammaticalizes does not have to be a lexical verb of speech; an expression of non-verbal origin ('as, like'), or a verb that does not literally code speech but rather motion or manner ('go', 'do so') will do as well, as long as it is conventionally associated with the report of a quote (and hence, as Deutscher states, "means roughly 'X said'"). Second, Deutscher's term 'reduction' refers to the fact that the unit 'X says' that grammaticalizes may start out as an entire clause, which becomes reduced to only its core element ('say', 'like', 'go') before this is reanalysed as a complementizer. Klamer (2000) shows that some languages provide an ideal morphosyntactic environment for this reduction process, in that (i) they allow subject indexes or NPs to be left unexpressed ('prop-drop'), (ii) the grammaticalizing verb 'say' is intransitive in the first place, so that it can constitute a clause of minimal syntactic complexity. A common configuration in discourse, then, is that the verb introducing the quote occurs on its own, in between the matrix verb of speech and the quote ('I spoke, <u>said</u>, "[quote]"'). In this form and position, it is thus easily reinterpreted as a morpheme that links the two clauses.[8]

In my sample, complementizers that have followed this pathway in one way or another are numerous. Relevant languages in the network include Abun, Barasano, Choctaw, Hausa, Hmong Njua, Japanese, Kana, Korean, Lao, Lezgian, Malayalam, Ndyuka, Noon, Santali and Tukang Besi. In the great majority of cases, the marker

---

[8] This is also confirmed by Deutscher's (2000) study of the Akkadian complementizer *umma*: This is a largely invariant non-verbal expression (and hence not a verb that could entail a more elaborate clause), and the speaker NP that it obligatorily co-occurred with at earlier historical stages came to be lost at some time before the Middle Babylonian period (Deutscher 2000: 79). This sealed its fate for subsequent grammaticalization as a complementizer.

hearkens back to a neutral verb of speech ('say')[9], which may also surface in a non-finite form. Thus in Lezgian, the Aorist Converbal form of the verb *luhun* has come to be a marker of indirect speech (cf. (218) below); as was mentioned in Chapter 4, a common persistence effect is that the marker is optional or disallowed when the CTP is the source verb from which it grammaticalized; this is also true of Lezgian, where *lahana* does not normally co-occur with the CTP *luhun*, while other utterance verbs do:

(218) Lezgian (Nakh-Daghestanian: Azerbaijan, Russia; Haspelmath 1993: 367)

Gada-di [*wič* *k'wal-e* *amuq'-da* *laha-na*] *haraj-na.*
boy.ERG self house-INESS stay-FUT **say-AOR.CONV** shout-AOR
'The boy shouted that (lit. having said) he would stay at home.'

Similarly, the complementizer *ennə* in Malayalam is the Perfective Participle of the verb 'say', and its reanalysis as a quotative complementizer "is found in all subgroups [of Dravidian] and is reconstructible for Proto-Dravidian" (Krishnamurti 2003: 451). In Barasano, a quote is followed by the Infinitival form *yi* of the verb 'say', which is in turn succeeded by a matrix clause that contains a semantically more specific utterance verb ('talk, cry, tell, ask'). Jones and Jones (1991: 61) are inclined to analyse the sequence of Infinitive and matrix verb as a "compound", e.g. *yi goti* 'saying-spoke'.[10] In Santali, the complementizer *mɛnte* is historically a combination of the verb 'say' and the Instrumental case suffix *-te*; literally, then, this amounts to 'he spoke/said, <u>by</u> saying [Quote]', which is not unlike the participial/converbal forms encountered above:

(219) Santali (Austro-Asiatic, Munda: India; Neukom 2001: 183)

Am-də-m *baḍae-kid-iŋ-a* [*iŋ-də-ɲ* *maraŋ-ge-a* *mɛnte*].
you.SG-TOP-2SG.SBJ know-PST.ACT-1SG.OBJ-IND 1SG-TOP-1SG.SBJ big-FOC-IND COMP
'You knew that I am the big one.'

And so on. A slightly different case of 'reduction' in Deutscher's sense is found in Korean. Here, the quotative construction is marked by the "quotative particle *ko* 'that'" (Sohn 1994: 56), which was illustrated in (188a) in Chapter 6. According to Sohn (1994: 11), the quotative particle "derived probably from" the more complex expression *ha-ko* (lit. 'say and'), "with *ha* deleted". Finally, it should be pointed out that quite a few constructions in cluster C are quotative complements par excellence, but they have not developed along the lines sketched above. In particular, they do not involve any grammaticalized quotative marker, but simply consist of an unmarked, structurally balanced clause that is attached to the relevant utterance predicates. Depending on the transitivity of the CPTs in question, one may say that such complements are directly inserted into an argument slot of the matrix verb, or that they are adjoined to the matrix clause. In usage-based linguistics, the cognitive

---

[9] In keeping with Güldemann's (2008) findings, some African languages in the sample are such that the complementizer of an utterance complement is likely to derive from a manner expression, such as an adverb 'thus, so' (e.g. Khwe, Krongo, Mayogo, Noon).

[10] There are further languages in which the description or glossing suggests that the complementizer is, in fact, a phonetic constituent of the main clause (e.g. Abun, Ndyuka, Noon), at least in those cases where it synchronically relates to an utterance verb (with other CTPs, it may be conceived as part of the complement).

operations underlying these two scenarios been described as 'superimposition' and 'juxtaposition', respectively (Dąbrowska and Lieven 2005).[11] In cluster C, quotative complements of this type are found in Mapudungun3, Tariana3, Tamashek3, Motuna3, Jamsay1, Hup2, Kewa or Awa Pit1.[12]

It is well known that, once grammaticalized, quotative complements can diffuse quite widely, acting not only as complement clauses but even in the adverbial domain (cf. Saxena 1995). The constructions in my data that contain a quotative subordinator but are found in a cluster other than C show an unusual extension in their CTP profiles. In particular, quotative constructions from Kana, Hausa, Hmong Njua, Khwe, Ndyuka and Malayalam end up in cluster B. They can extend into the perception, desiderative and even subject-clause domain, such as A-clauses in Kana and Hausa (cf. Chapter 5). This sets them apart from the typical quotatives in C, and moves them closer to constructions with an entirely different diachronic background, which we turn to next. The highly interesting fact about cluster B is that the majority of the complements it contains (except for the quotative complements and a few others) shows significant overlaps with **relative clauses** (henceforth RCs). In §5.4, we saw examples of a pathway from RCs to CCs in connection with the position of complement clauses, but the synchronic and diachronic connection between relative and complement clauses needs some further elaboration. This is because the constructions in cluster B constitute relative clauses of very different types, such as externally-headed, internally-headed and headless RCs, and the diachronic relationships to complementation are quite diverse and must not all be binned into an 'RC pathway'. Furthermore, we have to bear in mind that even where this pathway is highly likely, the resulting CC constructions may also end up in other clusters of the network, as they can co-occur with different kinds of CTP profile. Thus we also find a sizeable number of RC sources in clusters E and I, and scattered instances all over the network. These will be included in the discussion of historical issues, but I will comment on them separately where appropriate.

The development of pre-existing RC constructions into complements has been noted widely in the literature, most recently in Hendery's (2012) diachronic typology of relative clauses. She shows that for a number of languages, the CC usage of a particular construction is attested in historical documents only well after its widespread use as an RC. She argues for this to be the case in Hungarian (cluster B), Georgian (E) and a number of languages that are not included in my sample (e.g. Greek and Hebrew). Additional evidence to this effect has been reported for German

---

[11] Dąbrowska and Lieven are, of course, not concerned with quotative complements, but with the usage-based mechanisms that generally allow for the creation of complex symbolic units from previously experienced material. This idea is developed to account for data from first language acquisition, but the processes themselves are claimed to be fundamental to language production more generally. Consequently, they could also be expected to be at work in the emergence of certain types of complex sentences.

[12] These constructions are called 'quotative' here because they are either explicitly described as such in the sources or because they are complements that arguably took their starting point in quotation because their CTP profile involves primarily direct speech and thought but has become extended to selected other environments. The unmarked superimposition or juxtaposition of a fully sentential structure is also found in quite a few other instances in cluster C (e.g. Tetun1, Gooniyandi1), but here the evidence for a quotative origin is less clear.

(cluster B; e.g. Axel 2009) and Indo-European more generally (Lühr 2008). For Persian (cluster E), Öhl and Korn (2006) argue that the present-day complementizer *ke* was directly derived from its use in relative clauses, although it had had a long history before that: It ultimately goes back to the Old Iranian interrogative pronouns in the *ka*-stem, whose specific form *kahya* first gave rise to the relative pronoun *kē*. Thus in Middle Persian, relativization was rendered by *kē* while complementation was taken care of by the complementizer *kū*, so the two domains were coded distinctly. However, *kē* came to be reanalysed as a complementizer, competing with and ultimately ousting *kū* in this function.[13] Similar reanalyses from relative to complement clauses are also extremely likely for further constructions in cluster B, such as Burmese2, To'aba'ita1, Martuthunira1, Tamashek2, Purépecha1, Korean2 and Ainu, as well as constructions from other clusters (Japanese1 (I) or Newari2 (C)).

As discussed earlier in this dissertation, the crucial difference between RCs and CCs is that the former typically involve what Lehmann (1984) calls 'nucleus formation', i.e. the concentration on one of the arguments of the subordinate proposition. This is either extracted to become the external head of the RC, or it is left inside the RC ('internally-headed'). As Langacker (1991: 423) puts it, "the common feature of relative and complement clause constructions is that a main-clause participant is in each case coded by a constituent which either consists of or prominently includes a clause-like structure. The ultimate basis for distinguishing between relative and complement clauses is a semantic factor, namely […] whether the main-clause element in question corresponds to a *participant* of the subordinate clause, or to the *overall process* designated by that clause" [emphasis in original]. The question for a diachronic transition from relative to complement clauses is thus how this participant-designation can be reinterpreted as a processual one. As far as I can see, there are at least three contextual environments which may bring this about, and they are not mutually exclusive but may be present simultaneously in any given language. These possible bridging contexts are set out in Table 22 and discussed in turn below:

Table 22. Bridging contexts from relativization to complementation

Appositive environment	Product environment	Matrix-coding environment
I know the fact [REL you told me]. [+WH] 'I know the fact which you told me.'	I know the thing [REL you bought]. [+WH] 'I know the thing which you bought.'	I saw the woman [REL leave]. [+WH] 'I saw the woman who was leaving.'
I know the fact [REL she kissed him]. [-WH] 'I know that fact that she kissed him.'		
Antecedent drop: I know [REL she kissed him].	Antecedent drop: I know [REL you bought]. [+WH] 'I know what you bought.'	
Reanalysis of REL > COMP I know [COMP she kissed him]. 'I know that she kissed him.'	Reanalysis of REL > COMP I know [COMP you bought]. [-WH] 'I know that you bought (it/something).'	Reanalysis of REL > COMP: I saw [the woman leave]. 'I saw the woman leaving.'

---

[13] Interestingly, Öhl and Korn also mention that other Iranian languages, like Kurdish, developed differently: the equivalent of *kū* continued to be used as a complementizer and was not replaced by a grammaticalized form of *kē*. And note, finally, that Middle Persian *kū* had gone through the same development from relative marker to complementizer at yet older stages of the language.

The first environment was exemplified in (123) of Chapter 5: it involves the relative clause being generalized to what Comrie (1996) has called a 'noun-modifying' construction, or what is also known as 'fact-S', 'noun complement clause', 'appositive clause' or 'explicative clause'. Such constructions differ from relative clauses proper in that the head noun is not an argument of the subordinate proposition; instead, the whole proposition spells out the content of the head, which is itself an abstract noun such as 'idea', 'fact' or what some people call a 'light noun' such as 'thing'. If a language allows extending its relative-clause morphology to the expression of such appositive relationships, the function of the relativizer involved goes beyond that of a role-recovery device – it is generalized to a subordinator (cf. Hendery 2012).[14] All that is needed now to proceed from a nominal complement to a verbal one is the omission of the antecedent noun, which is semantically redundant, anyway, and hence prone to reduction. In this way, the subordinate clause becomes directly associated with the verb and can thus be reanalysed as a complement clause, and the relative-clause morphology ends up marking the complement. Axel (2009) argues that this pathway accounts for the development of German *dass*-complements. It appears to fit the diachronic data better than the commonly assumed scenario of clause integration whereby the complementizer derives directly from a cataphoric demonstrative in the matrix clause ('I know that: he kissed her' > 'I know [that he kissed her]').[15] A similarly appositive context was also one of the sources by which the relative-clause marker *asher* > *she-* came to be transmitted to complement clauses in Biblical Hebrew (Givón 1991: 291). But note that, more generally, it is not necessary for the antecedent to be omitted in order to derive a complement from an appositive structure. It is also possible for the antecedent itself to be reinterpreted as the complementizer. This happened, for example, in Dolakha Newar. An example of the resulting complement was given as (11) in Chapter 2 and is repeated here for convenience:

(220) Dolakha Newar (Sino-Tibetan, Tibeto-Burman, Bodic: Nepal; Genetti 2006: 145)

Āpsin	[ām	jāl	ta-en	ta-e	khã]	ma-si-u.
3PL.ERG	that	net	put-PTCP	put-NMLZ	COMP	NEG-know-3PST.HAB

'They didn't know that the net had been put there.'

---

[14] Historically, relativizers in this general sense may, in fact, go back to genuine role-recovery devices such as relative pronouns. This is what Lehmann (1984: 393) calls a "decumulation" process which unpacks the erstwhile portmanteau expression of three functions (subordination, role recovery and attribution) and reduces them to the subordinating one. In Croft's (2000) classification of reanalysis processes, this would amount to so-called 'hyperanalysis', in which a previous semantic component of a construction is 'analysed out'.

[15] Incidentally, this pathway of clause integration has also come under attack from other, and diachronically different, directions. McConvell (2006), for example, argues for the Ngumpin-Yapa genus (a subgroup of the Pama-Nyungan family) that the cataphoric pathway is implausible because "it is not possible in these languages […] to refer forward to [propositions] as 'this' or 'that'" (McConvell 2006: 133); the only cataphoric link of that sort is a manner deictic expression similar to English *thus* (i.e. **kuja*), and this morpheme has, indeed, developed into a complementizer for quotative complements. But other demonstratives are likely to have been reanalysed as complementizers along a different pathway. McConvell suggests that the demonstratives in question used to be *anaphoric* (rather than cataphoric) elements (e.g. 'the aforementioned'), which then merged with a pre-existing complementizer or replaced this earlier element altogether.

As Genetti (2006: 145) states, "this construction is clearly derived from a nominal complement construction in which a nominalized dependent clause modifies a noun within a noun phrase." The modified noun *khã* 'matter, story, talk, news' was reanalysed as a complementizer and, in the process, fossilized in the present shape (i.e. it cannot occur in the plural or with other nominal morphology anymore). This is, in fact, a frequent pattern in Asian languages, known also from Ainu *hi* and *pe* 'thing, place, person' (Bugaeva 2012: 468; cluster B), Korean *-n kes* (with *kes* = 'thing'; Yap and Matthews 2008; cluster B), Japanese *koto* 'thing, fact' (Wrona 2005; extending from cluster E) or Burmese <u>*hta*</u>/*hma* (*ha* 'thing' fused with an (Ir)Realis marker *te./me.* of the relative clause, cf. Soe 1999: 299; cluster B). Another case of a head noun used as complementizer occurs in Somali (cluster E), where *ín* 'amount, extent' has "grammaticalized" and is now "semantically empty" (Saeed 1999: 224); however, it still syntactically "acts like the headword of a relative clause in that it attracts certain elements out of the verbal group, notably subject clitic pronouns […]." (ibid.: 224) A typical complement clause thus looks like in (221) below:

(221) Somali (Afro-Asiatic, Cushitic: Somalia; Saeed 1999: 225)

Wáxa la íi sheegay [*ínuu akhrínayéy*].
wáxa one me-to reported COMP-he was.reading
'I was told that he was reading it.'

Further examples of the relative-appositive pathway may include Tamashek2 (Heath 2005: 674ff; cluster B), Koyra Chiini1 (Heath 1999: 186, 282; cluster B) and To'aba'ita1 (Lichtenberk 2008: 981ff.; cluster B).

The second conceivable bridging context between relativization and complementation involves what I call a 'product' environment in Table 22 above. In these cases, the relativized NP expresses the product (in the widest sense) of the action designated by the subordinate clause, e.g. 'the shirt [which you bought]', 'the thing [which you said]'. The absence of an overt antecedent results in a so-called headless RC. The question is now again how such a participant-oriented construction can come to be reinterpreted as a processual one, and here Lühr (2008), among others, suggests that there is a straightforward inference from product to process readings, especially with perception, knowledge and utterance verbs: If I saw what you stole, I also very likely saw that you stole something; If I know what you said, it is implied that I know that you said something; and so on. In other words, the product reading of a headless RC typically implies the process reading, though the reverse is not necessarily true. Lühr makes this responsible for the transition of Hittite *kuit* and Vedic *yád* from a relative pronoun [+WH] to a 'that'-complementizer [-WH]. As Lühr points out, the somewhat more difficult part is to account for the fact that the RC is deprived of the relativized argument ('I know [REL you said ___]' = 'I know what you said'), whereas the processual reading does not contain such a gap ('I know [COMP you said something]' = 'I know that you said something'). This can be explained in two ways: First, many languages allow for the free omission of arguments, especially if they are of low referential status ('it', 'something'). In such cases, there is no difference in surface form between the headless-RC source and the complement target ('I heard [REL you said]' = 'I heard what you said/that you said it'). Second, where the structure of the

relative clause is such that it still leaves some overt representation of the relativized element, either in the form of a resumptive pronoun or as a verbal index, the subordinate clause takes (or closely approximates) the form that it would also take in complementation. This is illustrated by the following examples from To'aba'ita; (222a) exemplifies a headless RC, while (222b) provides a complement clause of the same type:

(222) To'aba'ita (Austronesian, Eastern Malayo-Polynesian, Oceanic: Solomon Islands; Lichtenberk 2008: 1079, 991)

    a. *Ta   wane si   tala'a-da*   [*na kere   'ani-a   botho   'eri*].
       some man 3SG.NEG match-3PL.OBJ REL 3PL.NFUT eat-3OBJ pig   that
       'Nobody will be able to match (those) who have eaten the pig.'

    b. *Nau   kwasi   thaito'oma-na*   [*n=o   mata'i*].
       1SG   1SG.NEG know-3OBJ   COMP=2SG.NFUT be.sick
       'I didn't know that you were sick.'

Hendery (2012) points out in relation to Hebrew that the resumptive character of its relative clauses may have been a crucial facilitator of a complement reanalysis. More generally, it was mentioned above that the different bridging contexts in Table 22 are not mutually exclusive, and that languages may host them simultaneously. This arguably holds for To'aba'ita, and the Asian light-head constructions from above are not only available in an appositive reading, but also in a product interpretation. Thus Soe (1999: 302) states explicitly that "in the absence of extra-linguistic contextual cues", the following example "is ambiguous between a product-like or event-like interpretation":

(223) Burmese (Sino-Tibetan, Tibeto-Burman, Burmese-Lolo: Myanmar; Soe 1999: 302)

    [*Thu pyo: <u>hta</u>*] <u>*kou*</u> *ma-cai'* *hpu:*.
    3   say   NMLZ OBJ NEG-like NEG
    'I don't like the fact that he said thus.' or: 'I don't like what he said.'

We saw above that the morpheme <u>hta</u> derives historically from a head noun 'thing' and element of a preceding clause. The reanalysis to a complementizer, or abstract nominalizer, could thus have been brought about by an appositive context or a product reading from which a processual interpretation arose.[16] The product/participant pathway is also a very likely candidate for an interesting construction in Sanuma (cluster I). Here, a specific kind of RC construction involves a left-adjoined clause

---

[16] Another interesting case that should be mentioned in this connection is that of the Japanese complementizer *no* (cluster I). This is commonly analysed in the same fashion as its counterpart *koto*, i.e. as an "abstract nominalizer" (e.g. Josephs 1976) that arose in a relative/appositive construction (cf. above). However, Wrona (2005, 2007) argues that the sources of the two markers need to be kept distinct: *koto* is, indeed, a former head noun that grammaticalized, but *no* was used as a relative marker, not a head noun, from Old Japanese onwards. Its ultimate origin is that of a Genitive morpheme, and Wrona shows that this marker came to be occasionally inserted between a relative clause and a head noun, "typically when the head noun is of a particular kind, e.g. nouns of content like *kokoro* 'intention'" (Wrona 2007: 12). In other words, genitive *no* came to be used in appositive-relative clauses, and Wrona argues that it was then "interpreted as a nominaliser" (ibid.), thus bridging the gap to complementation. Interestingly, the earliest uses of nominalizing *no* without a head noun are indeterminate between a headless relative (product) and a complement (process) reading, so again both the appositive and the product pathway may have contributed to the development of the complementizer.

terminated by an indefinite marker, followed by a relative marker and an appropriate anaphoric representation of the head noun in the matrix clause. The resulting structure looks like this:

(224) Sanuma (Yanomam: Brazil, Venezuela; Borgman 1990: 136)

    [*Saöma töpö to-ma kõ wi*], *ĩ a hu kite.*
    Sanuma 3PL well-CAUS return INDF REL 3SG come FUT
    'The one who makes the Sanuma people well again will come.'

This specific structure is commonly used with the general, inanimate third-person pronoun *te* as the anaphor in the main clause, with the relative clause denoting precisely a product in the above sense. Significantly, the exact same structure can be also be used as a processual complement; its use as a subject clause was illustrated by (101) in Chapter 5 ('Stealing is bad.'). And this is again because certain contexts are ambiguous between a headless-RC and a complement interpretation, e.g. '[you cook INDF] 3SG good' = 'What you are cooking is good' or 'Your cooking/that you are cooking is good'.

A variant of the product pathway is also found where a pre-existing relative clause is used without an overt head noun, but where this headless construction involves the addition of nominal morphology in order to create a referential expression. Thus in Wolaytta1 (cluster B), Lezgian3 (outlier of cluster F) and Malayalam2 (cluster I), the basic structure used in RCs is suffixed with a gender marker (Malayalam)[17], or a derivational nominalizer (Lezgian, Wolaytta)[18]. This yields a headless RC or 'product' interpretation, but the same structure is also available as a complement clause with a processual reading. For relevant examples, please consider (40) and (93) in Chapter 4 again.[19] Later on, we will see that the product > process reanalysis, as proposed by Lühr, is, in fact, very common in nominalization constructions, and these need not involve a pre-existing relative clause. At present, we simply note that the product pathway constitutes one bridging context from relative to complement clauses.

The final bridging context from relative to complement clauses was dealt with in Chapter 4: We saw that in so-called 'pivotal' constructions, the subject of the

---

[17] In Malayalam, the structure consists synchronically of the tense-marked stem of the verb and the neuter gender suffix *-atə*. However, this is indistinguishable from the Relative Participle in *-a* with the gender marker *-atə* attached (and phonologically fused), so there is a possibility that this is the ultimate source of this headless RC construction. Hence Malayalam was tentatively included here.

[18] In Lezgian, this is a deadjectival nominalizer ('green' > 'the green ones') added to the Relative Participles (Haspelmath 1993: 110), and in Wolaytta, it is the more general nominalizer *-ga* that can also, for example, create a possessive pronoun from a possessive modifier ('my' > 'my-NMLZ' = 'my ones/mine', cf. Wakasa 2008: 87); this is added to verb forms that Wakasa (2008: 465ff.) argues to be basically relative-clause forms, so the complement would seem to be derived from them (although no direct diachronic evidence is available).

[19] An RC>CC pathway of a similar type may also underlie the development of finite complements in Basque (cluster B). Finite subordinate clauses in *-en* are likely to be of RC origin, as Trask (1997: 247) states that RCs themselves were "perhaps obtained by adding the genitive *-en* to finite clauses." The complement use of *-en* is said to involve "nominalized genitives" (ibid.), which basically refers to headless RCs. Finally, the complement-clause variant in *-ela* (the default declarative complementizer) "is widely thought" to have been "derived from *-(e)n* by the addition of a further morph *-la*" (ibid.: 240), which forms manner adverbs and related expressions. In other words, finite complement clauses in Basque historically derived in "a manner which is not entirely clear" yet (ibid.: 247), but which is likely to have to have its ultimate origins in relativization with *-en*.

complement clause is coreferential with the object participant of the matrix clause; consequently, it typically comes to be left implicit in the complement clause (= raising) or is repeated in pronominal form (= prolepsis). Both patterns involve 'matrix coding' (Van Valin and LaPolla 1997: 561) of the complement subject as the matrix object, hence the name of the bridging context in Table 22. As was argued in Chapter 4, the most natural environment is that of direct perception, where such a coding pattern is semantically motivated. When this is extended to other predicate classes, a genuine complement analysis may come about, involving either raising (e.g. 'I want him to go') or an even more general pattern of non-canonical subject marking in the complement. In Chapter 4, this was discussed in relation to Lithuanian and Kayardild ('Complementizing Case'), and a similar pivotal history also underlies certain complementation patterns in Martuthunira. (255) below illustrates the so-called 'Present Relative clause' (Dench 1995: 255; cluster B) in complement function:

(225) Martuthunira (Australian, Pama-Nyungan: Australia; Dench 1995: 255)

Nhuwana nhuura nganaju yilangu karri-nyila-a purnumpuru.
2PL   know   1SG.ACC here   stand-PRS.REL-ACC   quiet
'You know that I'm standing here quietly.'

In West Greenlandic, the so-called 'Participial mood' occurs in relative clauses but is also found in complementation. As in the other examples above, perception contexts are such that the subject of the complement is anticipated as the object of the perception verb:

(226) West Greenlandic (Eskimo-Aleut: Greenland; Fortescue 1984: 38)

Taku-aa [*aalla-lir-suq*].
see-3SG :3SG.IND   leave-begin-3SG.PTCP
'He saw him leaving.'

As Fortescue (ibid.) comments, "with verbs of seeing or 'coming across', the object inflection of the main verb may agree with the subject of the subordinate verb, but this is no longer the case with the other verbs, which are usually analysable as taking the whole subordinate clause as their object." In other words, indexation in perception contexts (here: 3SG:3SG) seems to be structurally ambiguous between pointing to a third-person subject of the complement clause and the complement as a whole. This, in turn, may have been the decisive bridging context to transfer the Participial morphology to complementation more widely.

This concludes our little survey of the major contexts in which externally-headed and headless relative clauses can become reanalysed as complement clauses with processual readings. What we have not yet touched upon are instances in which complements show overlaps with circumnominal relative clauses, i.e. those that have the head inside the RC.[20] Syntactically, circumnominal RCs can be directly embedded in the main clause or adjoined to a coreferential pronoun in the matrix clause; the former type is known as internally-headed RC, while the latter type is often (but not

---

[20] This term is adopted from Lehmann (1984, 1986), but in contrast to him, I am employing it as a hypernym for all kinds of RCs that contain the head noun, as will be seen shortly.

always) associated with a particular discourse-pragmatic profile and this specific functional subtype is also known as 'correlative clause' (cf. Andrews 2007b: 217). Complement clauses that are virtually identical to either of the two types of circumnominal RC are found in cluster B (Amele1, Lakota1, Wappo2, Korafe2) and elsewhere (e.g. Slave1 and Choctaw 1 in cluster I, Wambaya1 in cluster C). (227) provides an example from Amele:

(227)   Amele (Trans-New Guinea, Madang: PNG; Roberts 1987: 51, 47)

    a.   [*Mala   mel   heje   on   eu*]   *ija   f-ig-a.*
        chicken   boy   illicit   take-3SG-REM.PST   DEM   1SG   see-1SG-HOD
        'I saw the chicken that the boy stole.'

    b.   [*Naus   uqa   uqa   na   ho   qo-i-a   eu*]   *ija   d-ug-a.*
        Naus   3SG   3SG   of   pig   hit-3SG-HOD   DEM   1SG   know-1SG-HOD
        'I know that Naus killed his pig.'

It has been pointed out in the literature that circumnominal RCs lack explicit expression of the attributive relationship between the head noun and its modifying clause, and they may additionally lack overt morphological means for nucleus formation, i.e. for signalling that one of the arguments of the dependent clause is to be interpreted as being coreferential with the entire clause (cf. Lehmann 1984: 177ff. for detailed discussion). This holds for Amele above, although a conventionalized strategy for identifying the head is to move it to the beginning of the dependent clause; and in Choctaw, the absence of any nucleus formation leads to regular ambiguity of the surface structure, with the correct interpretation arising only from the context (cf. Broadwell 2006: 50). For this reason, it has been argued that circumnominal RCs without any nucleus formation are, in fact, very weak representatives of the category 'relative clause'; they are clausal nominalizations that normally act as complement clauses but which, under specific circumstances, are reinterpreted as relative clauses (Culy 1990, Lehmann 1984: 180); thus specific matrix verbs like 'know', 'remember', 'see', 'be good', etc. can be directed at a participant or a situation, and it is precisely in these contexts that implicit nucleus formation can take place, yielding a relative clause. Not surprisingly, the constructions in our network which show overlaps between circumnominal RCs and complement clauses share precisely these contexts in their CTP profile. Analogical extension then leads to distributional dissimilarities and different clustering in the network, but the core CTPs that act as bridges between IHRC and complement readings are firmly attested in all of them. Given that the relative-clause reading only arises under specific circumstances and is hence "the marked case" (Culy 1990: 206), the logical step would seem to argue that "languages will not use nominalized sentences as IHRCs without also using them as complements." (ibid.) In other words, the diachronic direction between relative and complement clauses is reversed for at least those IHRCs that do not involve nucleus formation (cf. also Lehmann 1984: 383), and I suspect that this even extends to some cases where such strategies are applied – post-hoc, as it were, to signal deviation from the default complement interpretation.

Having said that, no claim can, of course, be made that the historical directionality between complement clauses and structurally identical circumnominal clauses is always the one just described; in most cases, we can do no more than observe the synchronic pattern of multifunctionality. For the sake of completeness, let me consider two more of the above-mentioned cases. In Slave, the general factive complement clause is marked by the complementizer *nį̂*; this morpheme is also used as lexical nominalizer to derive place names and, in keeping with this, it can be used as an internally-headed RC in a locative function. In this specific context, then, complement and relative clauses work identically (while they are marked distinctly elsewhere):

(228) Slave (Na-Dene, Athapaskan: Canada; Rice 1989: 1317, 1245)

    a. [*ʔeyi*   *ʔaħít'e*   *nį̂*]   *lą́*   *raxegha*   *gonezǫ.*
        place/there 1PL.be COMP   really 1PL.for   LOC.is.good
        'We really like the place where we're staying.'

    b. [*ʔelá k'ínaret'are* *ké* *kadį̂la*   *nį̂*]   *negháyeyidá* *yíle.*
        airplane   on   2SG.go:out COMP   1SG.saw.2SG   NEG
        'I did not see you get off the plane.'[21]

The Australian language Wambaya normally has externally-headed relative clauses but in certain contexts, the head noun is treated syntactically as belonging inside the RC (as indicated by the second-position placement of the Auxiliary complex, cf. Nordlinger 2006). Importantly, it is these contexts that also invite a complement interpretation because the matrix verb is one of direct perception again:

(229) Wambaya (Australian, West Barkly: Australia; Nordlinger 2006)

    Ilinga gin-a   [*galyurringi gi-n*   *bardbi*].
    hear 3SG.M.A-PST   water   3SG.S-PROG run
    'He heard the water (which was) running.'

In this case, it seems to be fairly clear that the unmarked, paratactic complement structure can be reinterpreted as an internally-headed RC; the latter is extremely marginal in Wambaya (and hence 'marked' in Culy's sense from above), while the former is well-established.

Summing up the previous remarks, a great number of complementation patterns in cluster B of our *NeighborNet* are closely related to relative-clause constructions of different types. Although the diachronic directionality between the two construction types may not always be the same, the bridging contexts involve similar matrix verbs, and it is because of this common core in the CTP distribution, and apparently fairly similar preferences in further lexical diffusion, that the relevant complementation

---

[21] This specific locative type of relative-complement overlap has interesting parallels in other languages. For example, Genetti et al. (2008) report for Dongwang Tibetan that the nominalizer -*sa* derives historically from a noun 'place, earth' but has come to be a bound nominalizer on the verb in oblique relative clauses (including locative relativization); crucially, the same structure can also evoke a complement interpretation with perception verbs ('[He him hit-*sa*] I see', which "could mean 'I saw (the place) he hit him', where the nominalized is a headless relative, or 'I saw him hit him', where the nominalized clause is a complement" (Genetti et al. 2008: 140)). And a similar pattern was mentioned in §6.2 for Basque, where perception and phasal verbs require the nominalization construction to be enhanced by a locative marker rather than an article.

patterns end up in the same synchronic cluster. What sets some constructions with CC/RC overlaps apart is that they show specific CTP extensions or restrictions not shared by cluster B. CC/RC constructions in the lower branches of cluster I, for example, show extensions into the phasal-P domain (Japanese1, Slave1, Malayalam2, Choctaw1), while those in cluster H display a limitation to perception verbs, some S-clauses and at best one more of our P-environments, if at all (Santali2, Tümpisa Shoshone 4). And finally, the relative-based complements in Georgian, Somali and Persian discussed earlier end up in an area which I highlighted (by grey shading) as 'E'. As can be seen in the network, this area is actually pretty much the central plane of the entire graph, which simply means that it does not have a distinctive CTP distribution at all, i.e. it cannot contribute to any of the more strongly profiled clusters reaching outwards at all ends. And this, in turn, is because the relevant constructions show lexical diffusion 'across the board', i.e. extension to predicates of many CTP classes. The other constructions in this area thus show a similar 'success story', and, interestingly, they have developed along a number of very different historical trajectories. In fact, their different backgrounds nicely foreshadow the remaining major diachronic pathways that need to be discussed in relation to the other clusters of the network: (i) Some constructions in area E simply consist of an **unmarked** sentential structure that is either superimposed on or juxtaposed to the matrix predicate; a subordinator can be inserted occasionally with some CTPs, but all environments can also normally occur completely unmarked. This is true of Mandarin (MCh), Karo Batak (Bat), Vietnamese (Vie), Rama (Ram1) and Begak Ida'an (Beg1). (ii) The sentential complement in Serbo-Croatian is likely to go back to a **purpose** clause, as was discussed in Chapter 5 (cf. especially fn. 24). The purposive pathway to complementation is widespread but agglomerates extremely in cluster G. (iii) The complements in Barbareño Chumash (Chu1), Chalcatongo Mixtec (Mxc) and Musqueam (Mus2) show different forms of **clausal nominalization**, which at least in Mixtec reverses the diachronic directionality between relative and complement clauses. In Mixtec and Musqueam, the nominalizing element in question is also available as a lexical nominalizer, and so **nominalization** processes will have to be looked at more generally.

We first turn briefly to cluster D, which is characterized by the fact that most of its complementation patterns consist of **unmarked** sentential structures, i.e. not involving a subordinator that could point to a specific source construction. These complements are often described as full sentences in superimposition or juxtaposition; they do not reach out into the subject-clause domain, and while their CTP distribution in object function basically shows the same cognition-utterance profile as patterns in cluster C, it is often punctuated by certain 'holes' that are filled by alternative strategies (e.g. different-subject desiderative and knowledge in Mekens, perception and knowledge in Slave2, perception in Ma'di1, etc.). Interestingly, the further one follows the individual branches in cluster D outwards, the more these unmarked complements turn into what we called 'pivotal' structures that are indicative of some degree of interlacing (cf. Chapter 4) with the matrix clause. Thus the data points Lakota2, Tukang Besi2, Trumai2, Supyire3 or To'aba'ita2 are all characterized by the obligatory absence of a

complementizer (which is characteristic of a type of complement they immediately contrast with in their respective languages), and a CTP profile that typically involves pivotal argument sharing such as causative, jussive and perception contexts.[22] This is illustrated below for Tukang Besi, where the quotative-based *kua*-complement in (230a) contrasts structurally with the pivotal unmarked variant in (230b):

(230) Tukang Besi (Austronesian, Western-Malayo Polynesian, Sulawesi: Indonesia; Donohue 1999b: 407, 404)

a. *No-potae-m(o)* [*kua no-motindo'u na amai*].
   3RL-say-PRFV COMP 3RL-thirsty NOM 3PL
   'They said that they were thirsty.'

b. *No-'ita-ko* ['*u-moro'u te tee*].
   3RL-see-2SG.OBJ 2SG.RL-drink CORE tea
   'They saw you drinking the tea.'

We now move on to the clusters on the lower left side of the network, viz. F and G. Cluster G is extremely biased towards complements that show synchronic identity with **purposive** constructions – the vast majority of the complements in question (at least 22 data points in G alone) show such traces. It has been established beyond doubt in the literature that adverbial clauses of purpose can become complement clauses in specific semantic and syntactic niches, from which they can extend along a principled pathway of lexical diffusion. This is laid out in detail in Haspelmath (1989), who proposes the following directionality of diffusion: purpose > irrealis-directive complements (i.e. manipulative and desiderative verbs) > irrealis-potential (i.e. modal and future-oriented evaluative verbs) > realis-non-factive (propositional-attitude and utterance verbs) > realis-factive (cognition verbs and some evaluative contexts) (ibid.: 298). The clustering of purposive constructions in the *NeighborNet* reflects this increasing degree of diffusion and bleaching of the purposive meaning nicely. Thus on the left-hand side of cluster G, we find branches reaching out of the network (e.g. Chumash2, Huallaga Quechua4, Santali4), and these are limited in their CTP distribution mainly to jussives, possibly causatives and only very few related predicates of the same general semantic group. In Santali, for example, the purpose clause in *jɛmɔn* can be used as a complementation strategy in jussive contexts, but not more widely as a complement (Neukom 2001: 196); in Barbareño Chumash, a purpose clause acts as the complement of jussive and manipulative verbs (including 'persuade', 'let', etc.), as in (231):

(231) Barbareño Chumash (Chumash: USA; Wash 2001: 91)

ʔI =ka =s-as^hunač [*ču s-am -saxtaklaʔ hi=hoʔ=1=xɨp*].
TOP=then=3-order COMP 3-INDEF-crack.open DEP=DIST=ART=rock
'He ordered that they crack open the rock.'

Constructions with shorter branches in cluster G tend to be more inclusive, also covering desiderative, phasal and causative and other environments; and, interestingly,

---

[22] The structure is different in two of the outliers in cluster D, namely Mrt3 (Marthutunira) and Eve2 (Evenki), which go back to a purposive construction rather than an unmarked sentential type of complement.

the further one moves to the right in the network, the more 'infinitive'-like (in a *functional* sense) the structures become, being firmly established in the complement domain and including, significantly, also subject clauses.[23] This holds for the Infinitives in German, Hungarian or Serbo-Croatian, which appear on the left-hand side of cluster F (where we generally find a particular gravitation to subject clauses, as will be shown later), and for some purpose-based complements even further to the right. Martuthunira3 and Evenki2, for example, are transparently derived from a purposive construction, but they end up as outliers in cluster D because they can even stretch into the propositional-attitude domain.[24] And as Haspelmath (1989) also argues, purposive constructions in complementation are not confined to specific structural configurations; while desententialized constructions ("infinitives" in a *formal* sense) are typical, the purposive trajectory is also open to more sentential constructions, as in Santali and Chumash above.

As the use of purpose clauses in complementation is so well-known, there is no need for further exemplification here (cf. also Verstraete 2008 and Schmidtke-Bode 2009 for more typological data). Instead, let me add a few theoretical remarks on the purpose-complement trajectory. First, from a usage-based perspective, which generally emphasizes local, exemplar-based patterns of analogical extension, it is interesting to ask how exactly the transition and diffusion process takes place. Los (2005) offers a detailed account of the rise and spread of the English *to*-Infinitive that can shed light on this issue. Specifically, her discussion nicely demonstrates the workings of at least four mechanisms: (i) direct reanalysis of specific exemplars; (ii) formal analogy; (iii) functional analogy; (iv) systemic analogy. As for (i), there are several verb classes which allowed a purpose adjunct in such a constellation that it is easily reinterpreted as a syntactic argument. A classic example, already discussed by Bock (1931), is the following one:

(232) Old English (Bock 1931: 158–159)

    a. *þæt Herodes secð þæt cild* [*to forspillene*].
       'that Herodes seeks the child in order to destroy it'

    b. *þæt Herodes secð* [*þæt cild to forspillene*].
       'that Herodes seeks to destroy the child'

As can be seen in (232a), 'seek' can take a theme argument as its direct object, and the *to*-Infinitive is adjoined to it as an adverbial clause of purpose. However, given the widespread use of OV word order, the object NP could also be analysed as the object of the purposive verb (with which it is coreferential, anyway), resulting in the rebracketing in (232b). This direct reanalysis of particular exemplars can explain how a

---

[23] As Haspelmath (1989: 298) points out, impersonal predicates like 'be possible' or 'be necessary' fall into his irrealis-potential class, which still shares some significant semantic features of purpose, and these may hence be the first environments for the occurrence of subject clauses, from where these can again spread analogically.

[24] For the sake of completeness, it should also be mentioned that 'infinitives' with purposive overlaps are also found in other areas of the network, such as cluster A (Lezgian, Imonda, possibly Lango and Koyra Chiini); these do not cluster with the purposive constructions in G because they are either restricted to same-subject contexts or, if applicable more widely, do not cover jussive verbs, which appears to be a crucial distributional property of cluster G.

to-marked clause can end up complementing verbs that do not license *to*-PPs as their objects, but only direct (i.e. theme) objects, including 'seek', 'want', etc. This is a crucial step in the extension of purpose clauses, since there is no direct formal model in simple sentences to support it. By contrast, the earliest environments which showed the *to*-Infinitive as their complement were such that a goal-PP is also one of their argument frames in simple sentences; those verbs included directive verbs like conatives ('try') and manipulatives like 'persuade', 'force', 'instruct', etc. In these environments, then, purposive *to*-Infinitives are *both* a formal and a semantic match for the relevant argument position, providing evidence for factors (ii) and (iii) above. The extension of the *to*-Infinitive to verbs without such prepositional objects, as in (232), might have come about by direct reanalysis, but it was certainly also supported by semantic and systemic analogies: First, the verbs in question (monotransitive 'seek', 'want', 'fear' but also ditransitive 'order', 'permit' or 'forbid') are semantically similar to the earlier conative ('try') contexts (they are all part of Haspelmath's irrealis-directive group); and second, Los argues that the reanalysis was fostered by the fact that the finite Subjunctive clause had already been well-established as a complement of irrealis-directive verbs. In other words, the *to*-Infinitive 'copied' the syntactic possibilities of a formally very different construction with similar semantics, and it ultimately actually ousted the Subjunctive clause altogether. This is what we may call a 'systemic parallel' that promotes analogical extension (labelled (iv) above).[25] In sum, then, "the distribution of the *to*-infinitive [as a complement] at first mirrored that of the *to*-PP. As such it occurred not only as purpose adjunct but also as GOAL-argument after conative verbs [...] and verbs of persuading and urging [...] Here it was in direct competition with the subjunctive clause [...], and at some point seems to have been reanalyzed [functionally] as a 'non-finite subjunctive'. This [shift] probably accounts for the fact that we begin to see *to*-infinitives appearing as THEME-argument of verbs with meanings like 'intend' and verbs of commanding and permitting." (Los 2005: 17–18)

The second theoretical comment relates to the degree of confidence by which we can assert that a given complement clause actually emerged from a pre-existing purpose clause, either by reanalysis or by analogical extension. There is little doubt about this historical direction in cases where we are dealing with a clearly purposive conjunction, but oftentimes, as Haspelmath (1989) himself points out, the purpose clause is nothing but a case-marked action noun, i.e. a nominalization with an appropriate allative, benefactive or dative flag (case marker or adposition).[26] This can

---

[25] Interestingly, Los (2005: Ch.8) also suggests another, even more abstract type of analogical development in which the Subjunctive clause may have acted as a model for the *to*-Infinitive: Los argues that *to* in the Infinitival construction changed its morphological status from a bound morpheme in Old English to a free morpheme in Middle English. Crucially, this rather unexpected change may receive a more plausible interpretation if one takes into account that the Subjunctive clause, too, changed (at around the same time) from synthetic to analytic modes of expression (i.e. now being realised by a modal, i.e. free word), and this may have acted as a model for the distributionally equivalent *to*-infinitive (Los 2005: 192). If correct, such developments would provide evidence for highly abstract types of analogy entertained by the speakers of a given language.

[26] This effectively means that a cyclic development needs to be posited, from nominalizations/complements to purpose clauses and back to into complementation, as has been noted, for example, by Disterheft (1980: 198) for Indo-European: "Former oblique abstracts [i.e. verbal nouns, KSB] are capable, paradoxically enough, of reentering

complicate the diachronic scenario if there are complement-taking verbs that select dative rather than direct objects to begin with. In such cases, a dative-marked nominalization simply replaces the dative-marked object NP of those verbs, and this can (at least theoretically) happen completely independently of whether or not the same kind of nominalization is also used as a purpose clause. The CTP that takes dative-marked complements may thus be a sufficient trigger for the further extension of this pattern, and the purpose clause may be of systemic support but not causally involved in the *emergence* of the complement in the first place. Needless to say, only diachronic data with extremely fine-grained resolution could yield the 'correct' trajectory, but I doubt that any such data are (or will be) available. Even for English, for example, Los' (2005: 17) claim that the "earliest function" of the *to*-Infinitive "was that of the purpose adjunct" must remain conjectural because it is based on indirect evidence from Gothic and from the form of the construction (i.e. its being based on a *to*-PP). Importantly, however, the earliest records available for English already attest to the *to*-Infinitive being used as complement of verbs that take prepositional objects with *to*, such as the conative and persuading verbs mentioned earlier. Therefore, if the purpose clause is really the ultimate starting point of the development, then the transition "must have taken place a good deal earlier than traditionally supposed, i.e. in prehistoric OE or even Primitive Germanic" (Los 2005: 18), so it is, in fact, not directly traceable. Similar problems are also encountered in other languages. In Wolaytta, for example, there is solid evidence that a Dative-marked Infinitive, which is also a purpose clause, has come to serve as an argument of CTPs which normally only take Absolutive objects, such as 'want'; however, while it is completely plausible that 'want' drew directly on the semantically similar purpose clause, one must also consider that there are Dative-taking verbs in the system to begin with, and these may have been an alternative source for the spread of the Infinitive. The same is true of Krongo and the Serbo-Croatian Infinitive (Wayles Browne, p.c.). In other cases, no such problems arise; in Wambaya, for instance, the CTPs that take a purposive complement are either intransitive or take Accusative objects, but no Dative NPs. Hence, the Dative-marked complement is very likely to have been derived directly from the purpose clause. In sum, the general point is that, for lack of the relevant data, it can sometimes be very difficult to establish the diachronic source of a complement because several models may have been available. It is possible, and indeed not unexpected in an exemplar-based theory of linguistic categorization, that individual CTPs draw on different source constructions. This was demonstrated persuasively by Givón's (1991) account of how relative-clause morphology spreads into the domain of complementation. He argues that there were three distinct environments for reanalysis, and that different CTPs were found in either one of these. He thus concludes that

> analogy in diachronic syntax most likely operates across relatively narrow inferential gaps. [… It is] a down-to-earth, highly concrete process. What at first glance may seem to be global, abstract analogy […] often turns out upon more

the nominal system. [...] The infinitive can then be used throughout the nominal paradigm as was the original nominal verbal abstract, thus completing a full cycle of noun-to-infinitive-to-noun."

detailed examination to involve much smaller, more concrete – and above all psychologically more plausible – intermediate steps. (Givón 1991: 258, 298)

This does *not* invalidate the more abstract types of analogy that we encountered in the spread of the *to*-Infinitive above, because local analogies in form and/or function, and contexts for direct reanalysis, were amply attested; these are likely to have brought about the critical changes, but more abstract systemic analogies may have had a facilitative effect. Finally, the situation described by Givón also makes us aware of how difficult it is to assign a possible diachronic pathway to some of the sample constructions that show multiple-domain overlaps. In Motuna, for instance, a certain type of complement can be identical in form to a purpose clause and so it is likely that there is a direct diachronic connection between them; however, the construction also shares significant features with relative clauses (just as in Biblical Hebrew), and so there is likely to be a fairly intricate story to the diachronic organization of the complex-sentence system.

Before we leave the terrain of purpose clauses for good, we should briefly expand the scope to look at complements derived from **adverbial** constructions more generally. While purpose is certainly the most prominent type of adverbial relationship to give rise to complements, other semantic kinds of adjunct clause can do so as well, or at least act as a complementation strategy. An interesting thing to observe now is that many of the constructions for which this holds true in my data end up in close vicinity in the *NeighborNet*, depending on the specific adverbial relation with which they overlap. In Chapter 5, we saw for Akkadian and Epena Pedee that complement clauses can arise from causal clauses ('I informed you because/that something has happened'). This has also been argued, by Mahamane Abdoulaye (2008, 2009, p.c.), for Hausa *dà*-clauses, which may have entered complementation with emotional predicates ('happy/sad/surprised' in either object or subject function). In both Epena Pedee and Hausa, the relevant construction has a bias towards subject clauses, while the P-domain is not or not very productively covered. For this reason, the constructions are found in the lower regions of cluster F, which contain complements with precisely such a subject-clause bias. In Hausa, *dà*-clauses likely emerged from PPs headed by *dà*, which came to be extended to temporal and causal clauses, and it is the causal reading that Abdoulaye takes to be a specific entry point into complementation. In other cases, the temporal context is the one that directly overlaps with the complement domain, and the connection is found in verbs of direct perception: 'I saw when/that you stole the money'. Although the absence of actual diachronic data makes it impossible to determine the direction of the extension (after all, it is logically possible that a temporal reading is inferred from a perception complement), it appears that in many cases, it is the complement-taking verb that avails itself of a looser, adjunctive type of clause combination, and that we are hence dealing with an adjunct > complement pathway. The relevant constructions are overwhelmingly found in cluster I of the network, and mostly as neighbouring branches in the upper half of I. This concerns Lez4, Men1, Cht2, Man, and Mat2 in the lowest branch of E. In Lezgian, the Aorist Converb is a construction for conveying chains of actions and sometimes also manner (Haspelmath 1993: 376); in complementation, its use is restricted to *k'an*

'want, must' and *akun* 'see' (ibid.: 369). Similarly, clause-chaining constructions are employed in both Menya and Choctaw, again most prominently for perception verbs. In Choctaw, the relevant clause-linkers are switch-reference markers that mean 'and' or 'and then' in coordinate and temporal chains; their complement use is illustrated in (233) below:

> (233) Choctaw (Muskogean: USA; Broadwell 2006: 275)
>
> *Bonnie-at* [*bookóshi'-mã*   *issi'*   *átta-na*]   *písa-tok.*
> Bonnie-NOM creek-DEM.ACC deer be.LGR-DS   see.NGR-PST
> 'Bonnie saw (that) the deer (was) at the creek.'

While the chaining construction is still a looser complementation strategy in Menya ('You said this and I heard you' = 'I heard you say this'), it is embedded as a genuine complement clause in Choctaw. In Mangarayi, an Irrealis mood prefix has come to act as the linker of a 'generalized subordinate clause' in all kinds of functions (cf. also Merlan 1981, Dixon 2002: 88 for this pattern). While it is thus difficult to reconstruct how the various types of subordinate functions have interacted diachronically, it is noteworthy that direct surface ambiguities can arise between temporal and perception-complement interpretations of the subordinate clause:

> (234) Mangarayi (Australian, Mangarrayi: Australia; Merlan 1982: 16)
>
> *Ña-yiri+wa-ni*   [(*w*)*a-ŋaḷa-gala+wu-b*].
> 2SG-see-PST.CONT   SUB-1INCL.PL-hang.up-PST.PNCT
> 'You saw that/when we hung it up.'

Exactly the same can happen in Urarina:

> (235) Urarina (isolate: Peru; Olawsky 2006: 775)
>
> [*Bute itɕa-a*   *noe=ne*]   *kwara-he-ɨrɨ-a=lɨ*   *katɕa-ɨrɨ.*
> boat make-3.DF Noah=SUB   see-CONT-PL-3.AF=REM.PST   man-PL
> 'The people were watching that Noah made a boat' or 'The people were watching when Noah made a boat.'

What distinguishes the constructions in Mangarayi and Urarina is that they are extremely limited as complements in Mangarayi, while they have spread widely in Urarina, being available as a general different-subject complement clause. Therefore, the Urarina construction is found in a very different place in the network, namely at the lower end of cluster C (discussed earlier). Again, it remains unclear whether the complement is the source or the target in these cases, but given that developments from adjoined to embedded structures (Hale 1975), or from hypotaxis to embedding (Hopper and Traugott 2003), are attested more widely (e.g. also from adjoined to embedded relative clauses (Lehmann 1984: 386f.)), it would not be too surprising if the complement is historically derivative of the looser adverbial construction. More detailed discussion of this issue is provided by Fleck (2006: 236ff.) in relation to Matsés, where we again observe adverbial constructions coming to serve as complements of specific predicates; in this case, the adverbial source is undeniable since the clauses transparently involve an 'Adverbializer' that "makes adverbs from verbs or adjectives" (ibid.). And these adverbial clauses now serve a variety of complement functions. One adverbial suffix, for example, is used for both temporal

and perception contexts, others are specialized for same-subject coreference and hence can be used with phasal or conative verbs. (236) below contrasts the adjunctive and the complement reading of the construction:

(236) Matsés (Panoan: Brazil, Peru; Fleck 2006: 239)

    a. [*Dektato-kin*]   *tapu-ø*   *tan-nu.*
       climb.up-while:SS   ladder-ABS   test-INT
       'I will test the ladder by climbing up.'

    b. [*Dektato-kin*]   *tan-nu.*
       climb.up-while:SS   test-INT
       'I will try to climb up.'

As Fleck notes, the subordinate clause has come to substitute the "O slot" of some predicates (such as transitive *tan* above), "while […] it still retains the properties of an adverbialized clause" (ibid.: 239). There is no doubt, then, that certain types of adverbial constructions are diachronic precursors to complement clauses.

We now finally turn to constructions involving **nominalization** processes of various kinds. As Deutscher (2009: 199) remarks, nominalization is "an unsung hero" in the evolution of complex sentences, and although his specific attention is with lexical nominalization, his statement is applicable more widely to all kinds of nominalization processes. I cannot by any means do justice to the amount of research that has been conducted on historical processes involving nominalizations; I will instead have to restrict myself to exemplifying some of the most important pertinent developments and their outcomes. Perhaps the most conspicuous type of nominalization is the **lexical** one surveyed in detail by Comrie and Thompson (1985|2007), i.e. where a dedicated derivational morpheme is used, crudely speaking, to turn a verb into a noun. Lexical nominalizations are extremely widespread in the complementation patterns of the sample (cf. Chapter 4 again for a detailed survey); in the *NeighborNet* graph, they are predominantly found in clusters F, I and A, and largely excluded from C and D.[27] In cluster F, we find some of the 'strong nominalizations' (i.e. those with a particularly high degree of de-/recategorization, cf. Chapter 4), such as those from Georgian, Gulf Arabic, Jamsay or To'aba'ita. Their high degree of nominality correlates with the fact that they are flexibly available in all the functions that simple NPs could cover, including especially S- and A-clauses (cf. Chapter 6 again for this correlation); cluster F is generally characterized by substantial coverage of the subject functions, and the more to the right or the bottom of F, the more the relevant constructions are biased towards subject clauses. That is, the constructions productively cover few if any of the object environments defined for the present study (though perhaps others) but are a means for expressing subjects, e.g. Tümpisa Shoshone2, Tzutujil3, Persian2, Matsés3, Gooniyandi2, Ungarinjin2, etc. (but note that not all of these involve nominalizations). Constructions in cluster I (and H) have already been identified as having perception-

---

[27] Lexical nominalizations are, of course, not in principle excluded from occurring with cognition-utterance verbs (clusters C and D), but these combinations typically involve a reconstrual of the overall constructional meaning (e.g. 'think' plus nominalization = 'think about something', 'know' plus nominalization = 'know how to do something'). Therefore, the specific constructional meaning intended by our CTP classes is not covered.

knowledge predicates as the core of their CTP profile, although extensions into the desiderative, phasal and subject-taking domain are also common here. The lexical nominalizations in this area thus join in with this general characterization (e.g. Barasano2, Tariana2, Ma'di2, Mapudungun1, West Greenlandic2, Wari'2, Tukang Besi3, Kolyma Yukaghir1, etc.). In cluster A, by contrast, an important characteristic of the CTP profile is a bias to same-subject environments and this, too, appears to attract lexical nominalizations (e.g. Matsés1, Urarina1, Yagua2, Noon1, Tamashek4, Huallaga Quechua1 or Lavukaleve). Below is an example from Tamashek, where the lexical nominalization is used in a phasal-P environment:

(237) Tamashek (Afro-Asiatic, Berber: Mali; Heath 2005: 685)

     À-jənna   i-ss-ə̀nta   [é-wet].
     SG-rain   3M.SG.SBJ-CAUS-begin   SG-hit.NMLZ
     'Rain began to strike (= fall).'

The verb that such lexical nominalizers attach to is usually still capable of exerting its predicational powers in order to attract a certain number of arguments, modifiers and other adjuncts, yielding a more complex, clause-like, expression rather than just a new word-formation (cf., for instance, Lehmann (1984: 53, 376f.) on this point in relation to the Turkish nominalizer -*DIK*). As we saw in detail in Chapter 4, the nominal function of the whole structure often constrains the form of internal elements to nominal expression formats (e.g. possessively coded arguments, adjectival modifiers, etc.). However, it has also been argued in the literature that such NP-like structures can become more verbal over time, in a scenario that Heine (2009b) calls 'clause expansion'. This is a "gradual process […] from fully nominal complements […] to fully clausal constructions […], with each new stage characterized by a decrease in nominal properties and an increase in verbal and clausal properties" (ibid.: 42). Although Heine's paper rests almost exclusively on comparative synchronic data and provides little evidence for constructions that have actually undergone the full extent of his proposed development, there is firm historical evidence for the general direction from nominal to verbal morphosyntax. In the best-described cases, this development involves the retention of the nominalizer on the verb (and thus stops short of Heine's full trajectory), but a shedding of nominal marking on most other elements of the complement. This can be observed in the diachronic dynamics underlying the English Gerund (e.g. *John('s) playing the national anthem well*), as laid out in detail by Fanego 2004. Fanego demonstrates that the verbalization of English action nouns headed by -*ung/-ing* took its course from late Middle English onwards, in a very incremental fashion. As in other examples of diachronic change discussed earlier, the specific morphosyntactic context was crucial in both the reanalysis and the spread of the construction. It can be shown, for example, that the type of Gerund first affected by verbalization processes is one that was not introduced by a determiner; by 1600, this type of Gerund had acquired a strong tendency for its internal object to be in canonical (rather than possessive) form. By contrast, verbal nouns including a determiner retained possessive object coding in 90% of all cases at the same historical time. In other words, the presence of a determiner in a given usage token of the verbal noun blocked the verbalization process, while its absence fostered it. Another crucial

prerequisite for the verbalization was the behaviour of internal modifiers, in three specific ways: (i) For some lexemes, there had not been a surface distinction between adjectival and adverbial uses to begin with, thus yielding potential surface ambiguities in individual tokens of the verbal noun; (ii) Some adverbs could be used as a noun modifier; this had been possible in Old English but became more frequent in Middle English. (iii) An abstract type of analogy may have been at work: The verb-particle construction which had arisen in Early Middle English also affected nominalizations (e.g. *the drawing up of something, his going down*, etc.) and, according to Donner (1986) and Fanego, created the impression that Gerunds can take postverbal adverbs as a general pattern. As a result, where contextual factors were conducive (no determiner, adverbial modification), the verbal noun was interpreted as a more clausal structure, and its internal object came to be coded in regular NP form.[28] Basically the same trajectory of clause expansion has also been invoked for nominalized complements in Basque. Thus Trask (1997: 215) argues that lexical nominalizations in Basque go back to "noun-forming suffixes mostly meaning 'duration', 'abundance' or 'activity'", which "would have been ordinary nouns at first, with no verbal properties, but they came to be reinterpreted as verb-forms, and hence acquired verbal characteristics, just like the gerund suffix -*ing* of English." That is, the original verbal noun split off into two distinct constructions, a verbal noun proper and a new Gerund, just like in English. A remnant of the nominal origin is still visible in the northern dialects of Basque, which have the internal object in the Genitive (cf. ibid: 244f.). Finally, verbalization may also be able to account for the fact that some nominalizations in the sample (e.g. in Noon and other African languages) are simultaneously available in more NP-like and more clausal variants. However, one must also keep in mind that the opposite trajectory, i.e. from verbal to more nominal internal structure, has also been reported on in the literature. Andronov (1996), for example, suggests that the Malayalam Verbal Noun in -*al* is behaving more and more like a noun, so that possessive rather than canonical coding of the internal subject is becoming the norm. Therefore, care has to be taken in interpreting the kind of synchronic evidence that appears to be indicative of Heine's 'clause expansion' scenario.

The lexical nominalizations we have focused on so far are what we may call 'action nouns'. Semantic typologies of nominalization processes have also referred to these as "situation nominalizations" (Gerner 2012: 824) or "activity/state nominalizations" (Comrie and Thompson 2007: 334), as these morphemes derive a nominal designating a process in the widest sense ('do' > 'doing'); they are thus semantically well-adapted to the processual character of complements. However, there is also a second semantic type of nominalization, referred to variously as "participant nominalization" (Gerner) or "argument nominalization" (Comrie/Thompson). This construction takes a verb as its input and derives an agent, instrument, manner, product or location ('do' > 'doer', etc.). There is evidence that such nominalizers, too, can occur in complement clauses,

---

[28] Where the verbal noun *did* contain a determiner, it was the specific type of determiner that influenced the probability of verbal coding: The possessor type ('my doing X') was faster at acquiring direct-object NPs than the article type ('the doing X'), presumably because the former "has much greater affinity with an ordinary clause" (Fanego 2004: 37). Therefore, the more nominal in appearance, the less likely the verbalization process.

although several different diachronic scenarios are conceivable here. For example, where an agent nominalizer such as Quechuan *-q* (e.g. *pishtaku-q* 'slaughter-er', cf. Weber 1989: 53) is also used in subject-extracting relative clauses ((238a)) and in perception complements with raising ((238b)), it is likely that this involves the 'matrix coding' RC > CC pathway from Table 22 above:

(238) Huallaga Quechua (Quechuan: Peru; Weber 1989: 116, 290)

    a. [*willa-ma-q-(ni:)*]   *runa*
        tell-1OBJ-NMLZ-(3POSS) man
        'the man who told me'

    b. *Maria*   *noqa-ta* *rika-ma-ra-n*   [*Pablu-pa* *mishi-n-ta*   *qara-yka-q-ta*].
        Mary   1-OBJ   see-1OBJ-PST-3 Paul-GEN cat-3POSS-OBJ   feed-IPFV-NMLZ-OBJ
        'Mary saw me feeding Paul's cat.'

On the other hand, where a product or instrument nominalizer is involved, a more direct reinterpretation along the product > process inferential line sketched earlier is possible. In Matsés, for instance, the use of an instrumental nominalizer after CTP 'want' can be ambiguous between a participant and a situation reading of the subordinate clause:

(239) Matsés (Panoan: Brazil, Peru; Fleck 2006: 232)

    [*Kuesban-ø* *kues-te*]   *bun-e-bi.*
     bat-ABS   kill-INS.NMLZ want-NPST-1.S
     'I want bat killers.' (here referring to shot shells) or:
     'I want bat killing' = 'I want to shoot at/the bats.'

In keeping with Lühr's (2008) argumentation cited above, the desire for 'bat killers' (i.e. shot shells for killing bats) normally implies a desire for 'bat killing', so that the nominalization is easily reinterpreted as a processual complement clause. Arguably, the fact that instrumental nominalizations may also be used as relative clauses (when apposed to an appropriate 'head' noun) has little to do with the complement clause: The only CTP taking the above construction is the extended intransitive verb *bun* 'want', and there is no evidence for matrix coding or any other trace of a headed RC that may have given rise to the complement in (239); instead, it is far more plausible that the complement arose directly from a participant > process reanalysis of the nominalization. Similar synchronic overlaps between participant and situation nominalizers are also found, for instance, in Barasano, Yagua, Tümpisa Shoshone, Musqueam, Tariana, Mapudungun, and quite a few other languages. An example from Yagua is provided below:

(240) Yagua (Peba-Yaguan: Peru; Payne and Payne 1990: 336)

    a. *Jiy-dáátya-rà*   [*ray-jṵtay-janu*]?
        2SG-know-INAN   1SG-say-NMLZ
        'Do you understand what I am saying?' ('product reading')

    b. *Jááryiy ray-yisą́ta*   [*vícha-janu* *jirya*   *rooriy-mu*].
        much 1SG-be.bothered   live-NMLZ   this   house-LOC
        'I am bothered to live in this house.' ('process reading')

Again, there is no evidence that a (headed) relative clause was involved in the development of the complement: All that is needed is a nominalization construction that can be interpreted flexibly as product and process. Furthermore, relative clauses in Yagua do not involve the nominalizer *-janu*, but entirely different morphology; therefore, it is unlikely that they played a vital role in the emergence of the *-janu*-complement.

Basically the same principles carry over to what was called **clausal nominalization** earlier (cf. Chapter 4), where a relatively sentential structure is flagged by a nominal marker at the clause boundary in order to construe the clause in a nominal function. This can be determiner (e.g. the definite article in Lakota, demonstrative determiners in Tepehua, Barbareño Chumash, Jamul Tiipay and Tzutujil), a case marker (e.g. Trumai) or a nominalizer as such (e.g. Amele, Chalcatongo Mixtec). The latter are sometimes nominalizers with variable scope properties, i.e. they are synchronically available as both lexical nominalizers in word formation and extended verbal nouns, but can also be used to nominalize an entire clause. In Amele, the nominalizer *-ec* normally attaches to a verb stem and replaces "the subject person-number and tense inflection on the verb" (Roberts 1987: 275); additionally, it can be employed as a clitic to turn a fully inflected direct quote into "the object of the following quote verb" (ibid.: 48):

(241) Amele (Trans-New Guinea, Madang: PNG; John Roberts, p.c.)

[*Ija*   *h-ug-en*   *ec*]   *ma-d-om.*
 1SG   come-1SG.SBJ-FUT   NMLZ   say-3SG.DO-1SG.SBJ.REM.PST
'I told him I would come.'

In Musqueam, similarly, the morpheme *s-* is also both a lexical and a clausal nominalizer, and the two uses are "no doubt historically identical" (Suttles 2004: 265). An interesting case of a versatile nominalizer comes from Chalcatongo Mixtec. Hollenbach (1995) argues for Mixtecan more generally that the highly polyfunctional marker *xa=* and its cognates have their origin in a pronominal (i.e. demonstrative) morpheme that was used primarily to create participant ('product') interpretations from clauses, e.g. 'that what he said', 'that which is cruel' = 'the cruelties'. These 'headless relative clauses' are still common in Chalcatongo Mixtec:

(242) Chalcatongo Mixtec (Oto-Manguean, Mixtecan: Mexico; Macaulay 1996: 160)

Rùʔù   kuní=rí   [*xa=xíǰaa*   *núu*].
1SG   want=1SG   COMP=be.located   face
'I want the first one.'

Hollenbach argues that these structures gave rise to both idiomatic expressions that have now become fixed in the lexicon as words (e.g. 'to lie' > 'a lie') and, crucially, to processual complements. Her story is identical to the one we saw above: The product reading after certain verbs such as 'say' ('*xa=*you said' = 'what you said') gave rise to a complement interpretation ('that you said (something/it)'), precisely because an internal object ('it/something') could generally be left out. This was the trigger for using *xa=* as the general complementizer, as in (243):

(243) Chalcatongo Mixtec (Oto-Manguean, Mixtecan: Mexico; Macaulay 1996: 153)

$X^w\tilde{a}$ *nì-kunì=ø*   *[xa=nà-šuk^{w}ĩ́ĩ =rí]*.
Juan   COMPL-want=3   COMP=ITER-turn=1
'Juan wanted me to go back home.'

Interestingly, Chalcatongo Mixtec also uses *xa=* as its relativizer in *headed* RCs, so it is tempting to posit an RC > CC pathway again. However, Hollenbach claims that the extension of *xa=* to headed RCs is either a parallel or a later development as compared to the complement clause, but not an earlier one. This claim is based on the fact that all varieties of Mixtec exhibit the complement use, but only some feature the headed-RC use. In other words, she maintains that headed RCs using *xa=* came about by putting the original *xa=*nominalization next to a noun; these two referential expressions in apposition were then reinterpreted as a 'head noun' modified by a subordinate clause, so that *xa=* was bleached to a mere subordinator. This paved the way for a system-wide extension of subordinating *xa=*, e.g. also to adverbial clauses. Incidentally, the apposition of lexical or clausal nominalizations to nouns, and their concomitant reinterpretation as relative clauses, is a common diachronic trajectory: "Der Relativsatz wird als Substantivsatz konstruiert und dem Bezugsnomen als Attribut vorangestellt. […] Der Subordinator [hat] in solchen Sprachen primär die Funktion, Substantivsätze zu bilden, und diese Sprachen [...] verwenden Substantivsätze als Relativsätze" (Lehmann 1984: 53, 167). Lehmann identifies this strategy for a wide range of languages (including Turkish, Quechua, Navaho and others), although it appears to be particularly prominent in the Americas (cf. the contributions to Comrie and Estrada-Fernández 2012). Importantly, such cases, just like some of the circumnominal constructions surveyed above, suggest that complement clauses are not directly derived from headed-RC constructions, even though these share the same or a similar overall structure.[29] Instead, nominalization is, indeed, the "unsung hero", in Deutscher's terms, of syntactic complexity.

Needless to say, not all techniques of clausal nominalization can be subsumed under the same diachronic account. For many complementizers that derive from determiners, for instance, it remains presently unclear how exactly they have come to mark different kinds of subordinate clauses (e.g. Tepehua or Chumash); it is likely that clausal nominalization plays an important role here, too, but the synchronic picture is more complicated than in languages with straightforward ambiguities between relative, headless relative and complement clauses.[30] And finally, it is also evident that

---

[29] Yap et al. 2011 (§4.1) show that clausal nominalizations in East Asian languages are also often apposed to head nouns, yielding new relative clauses. However, in these cases, the nominalizer was commonly itself derived from a light noun as the head of a previous RC construction (cf. Ainu, Japanese and Newari above). This is also logically possible, but cannot be traced, for the nominalizers discussed in the present section.

[30] In Barbareño Chumash, for instance, it is unquestionable that the general complementizer *hi* is a demonstrative pronoun by origin that has now developed into a very versatile clitic that marks all kinds of dependent elements in NPs (e.g. possessors), simple sentences (e.g. verbal arguments) and complex sentences (e.g. subordinate clauses). However, the internal structure of relative and complement clauses with *hi* is radically different: Both headed and headless RCs contain nominalized verb forms, while the complement is sentential in nature. This makes the diachronic scenario quite complicated. Interestingly, Wash (1999: 62) notes that "Oblique-relatives that are *not* nominalized have the same form as noun complements", i.e. appositive clauses. This may have been a possible bridging context between relativization and complementation, although the precise trajectory remains unclear.

nominalizations can be quite diverse as far as their specific CTP distributions are concerned. While a common core can often be identified (as suggested earlier), they can also develop particularly wide diffusion patterns if they come to function as the major complementation technique (e.g. Turkish1 in cluster B or Musqueam2 and Mixtec in cluster E); conversely, they may be restricted in unexpected ways, ending up as outliers in the *NeighborNet*. For example, the construction labelled Tzutujil3 at the very bottom of cluster F is restricted to evaluative-S complements and not found with any of the other CTP-classes defined for the present study. Morphologically, it involves clausal nominalization since it is flagged by the definite article *ja(r)*, which is also employed for both headed and headless relative clauses. The diachronic trajectory is unknown, but the synchronic CTP distribution is very peculiar as compared to constructions with similar patterns of RC marking and multifunctionality; therefore, the construction ends up being an outlier in Material 8. Similar remarks apply to its neighbour Ungarinjin2: This 'generalized subordinate clause' typical of Australian languages is used as a relative, headless relative, adverbial and complement clause. The complement use is apparently more marginal than the other functions, and the only genuine attestation is with evaluative-S predicates (cf. (244c) below), although object functions are not explicitly excluded.

(244) Ungarinjin (Australian, Wororan: Australia; Rumsey 1982: 153, 149, 150)

    a. [*Malga   njadmaṛa-ŋari*],   *mindjal   biriṇji.*
       dance   we.did-SUB   eat   they.did
      'While we danced, they ate.'

    b. *Bedja   aŋga   [aŋulowani-ŋari].*
      already he.went   I.feared.him-SUB
      'The one whom I feared has already gone.'

    c. *Wa wuḷal wuŋge   [ŋabun   goydj ŋaya-ŋari].*
      not nice   it(W.CLASS).is.not alcohol drink we.go-SUB
      'For us to drink grog is not nice.'

Again, the diachronic developments cannot be traced, and it is the peculiar CTP distribution that makes the construction end up in cluster F, unlike many of the relative and adverbial patterns found in B and H/I, respectively.

In sum, we have explored in this section the major diachronic pathways that lead to the development of complementation patterns from other grammatical resources. Those included quotative constructions, relative clauses of different kinds (including externally headed RCs, appositive constructions, and headless relatives), purpose and other adverbial clauses, the juxtaposition or superimposition of unmarked sentential units into the main clause, and nominalization processes of different sorts.[31] The *NeighborNet* analysis showed that constructions with similar types of subordinators

---

[31] Because of the many difficulties involved in assigning a diachronic source to each construction in the database, I will refrain from venturing any sort of quantification of the different pathways in the database. We can be fairly confident that the sources listed above cover the lion's share of the complementation patterns in the data, but it would be more than adventurous to say that so and so many constructions actually derived from a relative clause, for instance.

and constructional polysemies tend to share certain commonalities in their CTP profile, leading to clustering in the network. Assuming that similar subordinators and polysemy patterns are indicative of diachronic origins, one could conclude, indeed, that the source construction tends to influence the lexical diffusion of a given complement, and that there is a certain amount of cross-linguistic convergence in these diffusion processes. Of course, each complementation pattern goes its own unique way of establishing a functional profile, finding its niche in the grammatical ecology of the complementation system as a whole. Nevertheless, there appear to be certain systematic patterns of CTP attraction and 'repellence' (to borrow Stefanowitsch and Gries' (2003) corpus-linguistic terminology) that recur in similar ways across the sample. In the next section, I will thus try to turn the perspective around and ask which CTP classes tend to show similar behaviour of attraction and repellence, thus producing the specific shape of the network.

7.2.2 The cross-linguistic clustering of complementation environments

The analytical step to be performed in this section is quite straightforward: If each complementation pattern in the data shows a particular distribution over the various environments, then each environment ends up being characterized by a certain pattern of selection by the 228 complementation patterns. As a result, the CTP classes can be compared to each other as to how similar their selection patterns are, yielding a principled, data-driven classification of complement-taking environments. There are several statistical techniques available for pursuing this goal. Croft and Poole (2008), for instance, make a strong case for multidimensional scaling of typological data sets, and very similar in spirit is the family of cluster-analytical methods that has already been applied above and in other places of this study. We will thus return to the latter kind of statistical procedure here. Specifically, I am going to perform a Hierarchical Agglomerative Cluster Analysis (cf. Gries 2013: 336ff. for an introduction); given that the number of data points to be grouped is considerably smaller than before, comprising just the 17 CTP classes, the tree-like output of a HACA will be more readily interpretable now than it would have been in the previous cases, where we grouped 228 data points.

   As in all cluster-analytical methods, two steps are necessary: first, the co-occurrence patterns are transformed into an appropriate similarity metric; second, the resulting (dis)similarities between the CTP classes can be used to form statistical groupings, based on a so-called amalgamation rule (i.e. a specific mathematical algorithm according to which the grouping proceeds). It is important to note that while cluster-analytical procedures are used to detect the underlying statistical structure in the data, both of the methodological steps just described involve choices between a variety of alternative techniques (i.e. between different measures of similarity, and between different amalgamation rules for clustering the data). Consequently, the resulting solution actually *imposes* a particular structure on the data (cf. Aldenderfer and Blashfield 1984: 16 on this point). As Baayen (2008: 146) remarks, "there are no hard and fast criteria to help decide what kind of clustering is optimal for a given data set. […] Typically, the [solution] shown is the one that best fits the authors' hypothesis of

the data", after having "tried out a variety of clustering techniques." This is exactly the procedure I am following here. In keeping with what was said in §7.2.1 above, I settled on a similarity (rather than a distance) measure in order to capture the curvature of the CTP vectors. I then computed the cluster analysis using different amalgamation rules. In the end, all versions yielded similar results, but the version that I believe to represent the data best is one that uses the so-called 'cosine' as a similarity measure (i.e. the one used above), and 'ward' as an amalgamation rule, which sidesteps some potential problems of alternative clustering algorithms and "has proven quite useful in many applications" (Gries 2013: 347). The resulting cluster solution is presented in Fig. 16 below:
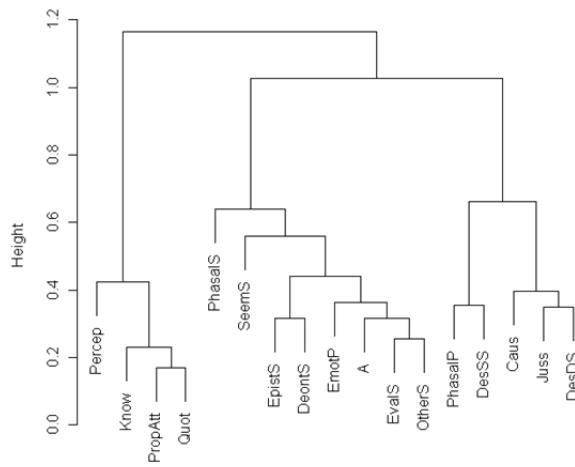
Figure 16. Dendrogram of CTP classes according to their distributional similarities[32]

Before we can discuss the groupings from a linguistic point of view, it is important to submit the solution to validation techniques. The idea is to determine mathematically the optimal number of clusters to distinguish in a given dendrogram, i.e. to detect the ones that are distinct enough to be considered relatively strong groupings. A commonly applied measure in this connection is the so-called 'average silhouette width' (*ASW*, cf. Kaufman and Rousseeuw 1990: 86ff.), which is essentially a goodness-of-fit measure for each possible 'cut' in a given cluster solution (e.g. a value for how well each data point lies within its cluster if we cut the dendrogram at N = 2 clusters). In this way, the *ASW*s for all possible numbers of cuts (here: N = 17-1 = 16) can be compared to each other, and higher *ASW*s indicate better cutting solutions.[33] As applied to Fig. 16, it turns out that the cluster achieves low *ASW*s on all possible cuts (according to Kaufman and Rousseeuw's (1990) standards), but it should be borne in

---

[32] The analysis was performed by using the `hclust` routine in *R*. For comparative purposes, I also fed the same similarity metric into *SplitsTree4* again. The resulting *NeighborNet* representation of the data is shown in Material 9 in the Appendix. As can be seen, the structure of the network is very similar. The minor differences result from a different amalgamation rule used in *SplitsTree4*, as noted in Chapter 4.

[33] In *R*, average silhouette widths can be calculated by using the function `silhouette` in the package `cluster`. I would like to thank Stefan Th. Gries for providing a number of additional commands for clustering and cluster validation.

mind that exactly comparable *ASW* values have also been interpreted meaningfully in other recent linguistic research (e.g. Gries 2012: 66). So, on the assumption that the *ASW* values can be taken to be indicative at all, the best partitioning of the data is found in the two-cluster solution (*ASW* = 15.9), followed by a four-cluster solution (*ASW* = 13.4). These solutions are displayed in Fig. 17:
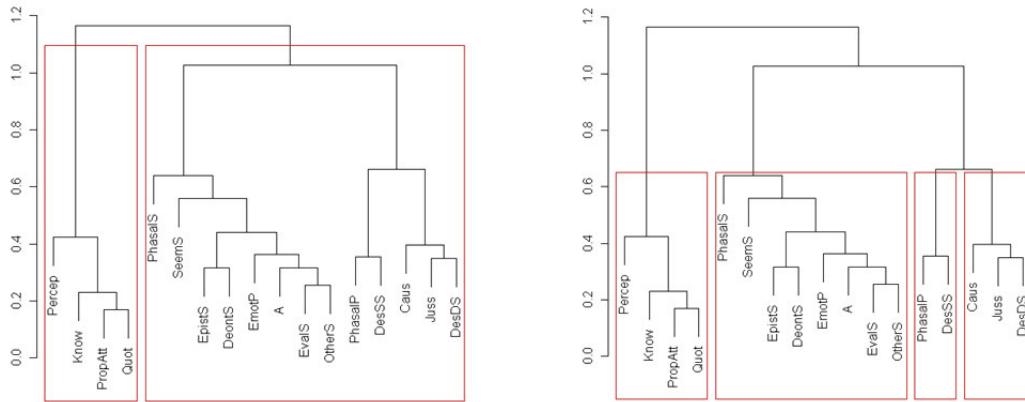
Figure 17. Best-fitting cluster solutions of the data (using 'ward' as amalgamation rule)

From a linguistic perspective, Fig. 17 captures similarities between different CTP classes according to how individual complementation patterns select for them across the sample. The clusters thus reflect points in the data where two or more CTP classes co-occur with the same construction, where they are repelled by the same construction because there is an alternative complementation pattern, and where they are repelled by the construction because the relevant functions are rendered outside of the domain of complementation, e.g. by monoclausal strategies. All of this information (in numerical form) entered the similarity calculation and hence needs to be taken into account when interpreting Fig. 17. In many respects, the results are, of course, similar to those offered in connection with the binding hierarchy of complementation, as discussed in Chapter 6, and hence are best interpreted in this light. For example, Fig. 17 (left panel) replicates a basic split between object clauses of cognition-utterance verbs (in Givón's sense) on the one hand, and subject clauses and modal and manipulative object-clauses on the other. Within the cognition-utterance group, there is a split again between propositional and predicational layers of clause structure: predicates of direct perception relate to predications, while all others in this group take propositions as their complements. In keeping with previous research (Ransom 1986, Dik and Hengeveld 1991, Cristofaro 2003), we saw in Chapter 6 that this difference is reflected in the preferred morphosyntactic structure of their complements. However, it must also be noted that Fig. 17 does not simply capture a form-function hierarchy. If it did, then perception verbs should cluster with jussive, causative and DS-want predicates: As our MDS scale in Table 20 of Chapter 6 showed, all of these predicate classes have similar structural preferences (e.g. for dependent verb forms and their overall CID). But the grouping in Fig. 17 is rather different: Perception verbs go their own way, so to speak, while causative, jussive and DS-want predicates clearly cluster together (they even form their own valid cluster in the second solution in Fig. 17). This

partitioning of the data reflects that an overall similar structural effect (e.g. morphosyntactic reduction or special argument coding) can come about in different diachronic ways: We saw in §7.2.1 above that an agglomeration of jussive, causative and (DS-)desiderative predicates is a typical diffusion pattern of complements that overlap with (and in many cases are historically derived from) purpose clauses. They share with purpose clauses their prospective cause-effect (or purpose-result) structure, and despite the fact that jussives and causatives, for example, differ as to whether they imply the realization of the intended effect, they are still prone to be targeted by the same complementation pattern (e.g. Chumash, Rama, Supyire and many other languages). Perception complements, by contrast, were argued above to draw on different historical sources, such as relative clauses, other types of adverbial constructions, lexical nominalizations and unmarked pivotal structures. [34] Of course, it is also perfectly possible for perception verbs to attract the same kind of structure that is used as a purpose clause: the latter is typically a control-inducing construction (cf. Schmidtke-Bode 2009: §3.3 for discussion), and these are suitable as perception complements to the extent that they can be construed as involving object control ('I saw him [ø leave]', cf. Chapter 4 again). However, in lexical diffusion, this syntactic similarity appears to be weaker than the semantic one uniting jussive, causative and desiderative verbs, so that perception verbs end up in a different cluster from the latter.

Syntactic similarities of argument sharing seem to be stronger for phasal-P and same-subject desiderative predicates: These two CTP classes appear as a distinct cluster on the right side of Fig. 17, and given that the time reference and factuality of the complement situation is different for phasal and SS-desiderative predicates, their similar distributional profile must to a large extent be due to the fact that both involve same-subject configurations.[35] They are thus likely to be covered together where languages have developed dedicated 'same-subject complements' (e.g. Lao, Urarina) or other structures with obligatory control of the complement subject (e.g. Mayogo, Noon, Wari', Lango, Quechua and numerous others). (Note that in many of the relevant cases, the construction is built around a lexical nominalizer and not a purposive subordinator. This, too, is what makes phasal-P and SS-want predicates different from the jussive, causative and DS-want classes.) What must also be noted in this connection is that phasal and SS-want predicates also cluster together because of their *not* being selected by any complementation pattern. As was noted in Chapter 6, the two environments are particularly susceptible to clause union (auxiliation, compounding, serialization) or morphologization. They encode 'secondary concepts' in Dixon's sense and show tight semantic integration between the matrix and the

---

[34] In Dixon's (2006a) typology of complementation, he also recognizes that both his so-called "ACTIVITY" type of complement (which includes direct perception contexts) and his "POTENTIAL" type of complement (which includes desiderative and jussive) tend to be characterized by reduced morphosyntactic structures, but different kinds of reductions in each case. This is a synchronic observation, but, as we have just seen, it can easily be related to the respective diachronic sources.

[35] Historical evidence for the influence of argument-structural similarities on lexical diffusion in complex sentences has been provided by Schulte (2007), who shows that the spread of Romance Infinitives in the adverbial domain is crucially influenced by the likelihood of each adverbial relation for same-subject constellations.

complement event. The latter is due to subject coreference in both cases, and additionally because of spatio-temporal inseparability in phasal-P verbs and 'agent binding' (in Givón's 1980 and Verstraete's 2008 sense) in SS-want verbs. They are thus rather weak candidates for fully biclausal expression types. As a result, both of them may grammaticalize along the same path (e.g. as "verb compounds" in Chimalapa Zoque (Johnson 2000: 117ff.)) or choose different routes of monoclausal development (e.g. desiderative affixes but phasal auxiliaries, as in Jamul Tiipay or Tepehua), but in either case, the result is located outside the domain of complementation as defined in the present study. Therefore, the remaining complementation patterns in the relevant languages 'sidestep' phasal and SS-want environments in their diffusion processes, resulting in the clustering of Fig. 17.[36]

What is finally detected by the cluster analysis is a group of environments that comprises all subject clauses (and emotive-P predicates); they even form a distinct cluster in the right version of Fig. 17. While it is true that most subject clauses share a certain proclivity for desententialized coding, the primary reason why they cluster together here is not to be sought in their morphosyntactic properties or preferred source constructions (such as nominalizations). Rather, they group together because of their low productivity or scarce attestation in the data, up to their collective absence in some of the sample languages. As was emphasized in Chapters 5 and 6, this absence must be taken with a large grain of salt, as it is likely to reflect gaps in the primary corpora and materials rather than genuine distributional restrictions. Therefore, future research will most certainly break up the present cluster and yield a more differentiated picture of how individual SA-clause-taking classes integrate with the remaining environments. At present, the internal grouping of the cluster mirrors almost perfectly the productivity gradation observed in Chapter 6 (compare with Table 20 again): phasal-S > seem-S > epistemic-S/deontic-S/A > other-S/evaluative-S.[37] In other words, where complementation patterns reach out into the subject domain, evaluative-S is the most likely candidate (on average) to be covered and, apparently, this often happens in conjunction with some 'other-S' environment. But again, given the rather shallow and undifferentiated cluster structure, little more can be said about the subject classes at this point. What is also an interesting avenue for future research in this connection is the historical pathways of emergence and analogical extension along which subject clauses develop, i.e. how they generally fit in with what has been said in the present

---

[36] An interesting precursor to the monoclausal development is the phenomenon that in some languages, such as Kato Batak, phasal-P and SS-want environments are the only ones in which a certain type of complement clause cannot be enhanced by a particular subordinator, while all others apparently can. This is indicative of the tighter structural integration of the two predicate classes with their complements, and the absence of a subordinator would seem to be a crucial prerequisite for processes like auxiliation or morphologization.

[37] The position of emotive-P predicates in the cluster does not make sense from this perspective since these have a markedly different productivity score (cf. Table 20 again). Apparently, there is some correlation to the effect that when certain types of subject clauses are not attested or ruled out, this also holds for the emotive predicates in question, or they are both covered, while other CTP classes behave differently. It seems to me that this is a largely accidental pattern, possibly due to the fact that the emotive-P class is not a semantically coherent group in the first place (cf. Chapter 2). I will thus desist from pondering possible motivations. (Note that the *NeighborNet* representation of the data in Material 9 places the emotive-P group slightly differently to begin with.)

chapter. The subject-taking CTP classes are semantically quite diverse, ranging from factive meanings in retrospective evaluation to epistemic and deontic modality, so there is, in principle, a broad spectrum of source constructions that would make an appropriate semantic fit. The difficulty lies in disentangling cases of emergence from those of analogical extension: Given that subject clauses are the typologically recessive type of complement, as has amply been shown in the present study, it is plausible to assume that many instances of subject clauses are analogical extensions from pre-existing object clauses. In Chapter 5, we encountered *kīma*-complement clauses in Akkadian, which arose by the reanalysis from a causal adverbial clause to an object clause. When consolidating as a grammatical object, such clauses later acquired the ability to passivize, and it is in this way that subject clauses came about (cf. (143) in Chapter 5 again). Crucially, this is the only environment for *kīma*-complements in the subject domain: "They are never found in A function, and only rarely in S function, always as a result of passivization of the main verb." (Deutscher 2006: 162) This is a rather clear case, therefore, of subject clauses being analogically extended, via passivization, from object clauses. Similarly, when complements emerge via a quotative pathway, it is rather uncontroversial that the subject uses of such quotative complements are later developments than the original use with utterance-P verbs. As we saw earlier, quotative complements often resist being used as subject clauses altogether, but diffusion processes can push them that far, as in Hmong Njua (cf. (120) in Chapter 5 again). But such uses are almost always diachronic offshoots of the corresponding object clause.[38] And in a similar vein, Haspelmath's (1989) paper discussed above proposes that the first environments in which purpose clauses are used as complements are predicates like 'order', 'ask', 'cause' and 'want', while the 'modal' predicates ('be necessary', 'be possible') are only in the second group of analogical extension; crucially, the first group has a clear bias towards object clauses, while the second one includes S-clauses. Here, too, then, an extension from object to subject clauses appears to be likely in many cases.[39]

However, in many other cases, we cannot be sure that subject-clause environments were populated by analogical extension. For nominalizations (both lexical and clausal), for example, it is perfectly conceivable that they can arise in subject function, although this is immensely difficult to trace in even fine-grained diachronic data; as Gerner (2012: 833) observes in relation to Asian languages, "many nominalizers […] have opaque sources", dating too far back in time to be discernible in their emergence and earliest usages. But it is not only nominalizations that may give rise to subject clauses.

---

[38] Exceptions to this general trend may occur where the quotative marker can also evoke causal interpretations in reported speech, as in Kana (cf. Ikoro 1996: 283). Since subject clauses can be derived directly from causal causes, there is at least a theoretical possibility that, for example, commentative (i.e. evaluative) S-clauses developed along this line.

[39] Haspelmath makes no claim as to the syntactic functions of the complement, so his proposal for analogical extension is cast entirely in semantic terms. It is thus perfectly possible that even the first stage involves a subject clause (e.g. where the notion of 'desire' is rendered as an experiencer-object pattern). However, this does simply not appear to be the typical case. And there are, after all, purposive complements which have not been found in the subject domain at all (e.g. Barbareño Chumash, Evenki, Huallaga Quechua, Fehan Tetun), thus strengthening the case for the 'object priority' of the purposive pathway.

Subject clauses may also be derived directly, via reanalysis, from a given source construction, and this may happen independently of (or even previously to) the development of object clauses from the same source. In Chapter 5 (example (142)), we saw how a sentence-initial causal clause in Epena Pedee may be interpreted as being the subject of the main clause, given that an alternative filler for the subject position is absent (or pro-dropped); if this is, indeed, an incipient complement clause, it is currently restricted to the subject domain – at least it has not been found as a regular means for rendering object complements. For Lao, we saw in example (139) of Chapter 5 that an unmarked complement clause can occupy the subject position directly or be left-adjoined when an explicit anaphor ('it') is found in the ensuing matrix. In light of the developments from adjoined to embedded structures that were discussed in the last section, it is thus possible that the embedded subject clause is simply a later, "tighter" (Enfield 2007: 467), variant of the adjoined structure, and hence developed quite independently of the corresponding object clause.[40] Explicit argumentation to this effect has also been provided by Wrona (2008) for a specific complement type in Old Japanese. Old Japanese had a particular morphological form of the verb called the 'Adnominal form', which is no longer existent but can be seen as the historical precursor to the modern *no*-complement construction that is also part of my sample. The Adnominal form had a fairly wide range of usages or functions in Old Japanese, and Wrona argues persuasively that the function implied by the name 'adnominal', i.e. a modifying or relative-clause function, was the primary syntactic context in which the Adnominal form occurred, and that "most non-relative usages are secondary or derived" (ibid.: 276), including sentential complements. As can be seen in Fig. 18 below, object clauses derived from headless variants of the Adnominal relative construction, and again the bridging context was most likely that of immediate perception; then they quickly extended to declarative-knowledge predicates (cf. Wrona 2008: 284).

---

[40] In this context, Fanego's (2004) study on the verbalization of the English Gerund is insightful again. Apart from the variants of the Gerund discussed earlier in this chapter, there is also one available in Present-Day English without *any* other nominal features (no possessive arguments at all, adverbial modification, no determiners, e.g. *John playing the national anthem well*). Fanego (2004: 44) argues that this type of Gerund developed independently of the verbalization processes of the other types we saw above. Specifically, she maintains that it emerged by the reanalysis of an Absolute Participle (e.g. *They$_i$ following the vein of the mine, ø$_i$ did dig forward still*) that came to be interpreted as the subject of the main clause. In other words, it followed a pathway from adjunction to embedding, just as discussed here for other languages.
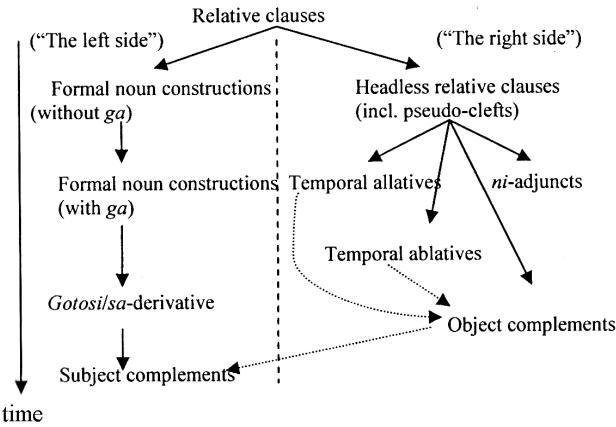
Figure 18. The development of subject and object clauses from the Adnominal (relative-clause) construction in Old Japanese (taken from Wrona 2008: 277)

Now, the most interesting fact about Fig. 18 is that subject complements are *not* a direct analogical offshoot of the object clause. Instead, subject clauses ultimately go back to the same original relative-clause usage of the construction, but do so via a chain of highly specific intermediate constructions (as can be seen on the left-hand side of the graph). There is no room and no need to present those constructions here, but the general point is that "subject complements and object complements developed differently, at least initially" (ibid.: 285). Again, the bridging context for the rise of the subject clause seems to have been that of 'commentative'/evaluative predicates. Once subject and object usages of the construction had established themselves in their respective niches, further interaction between them took place. Specifically,

> the dotted arrow from object complements to subject complements is meant to capture the fact that subject Adnominal complements of predicates other than commentative predicates are derived from transitive predicates through (medio-)passivization. In addition, it is to be expected that if the development of Adnominal subject and object complements did indeed take different developmental paths, then it is likely that a certain amount of analogical pressure was exerted on Adnominal subject complements, which developed slightly later than object Adnominal complements to all appearances. (Wrona 2008: 285)

In conclusion, even though it can be demonstrated that subject clauses arose only after object clauses and share their morphosyntactic structure with them, they were not all direct functional extensions of object clauses; instead, some of them developed from the same ultimate source construction, but via an independent pathway of reanalysis. It is to be hoped that further diachronic studies in Wrona's spirit will shed more light on the evolution of subject clauses.

7.2.3 Résumé

The preceding sections were concerned with patterns in the synchronic distribution of complement clauses over various complement-taking environments. These patterns arise through processes of lexical diffusion, which in turn are driven by the cognitive mechanism of analogy. In the historical-linguistics literature, lexical diffusion and

analogical extension have been treated as the 'actualization' of language change, i.e. as the language-internal propagation and implementation of a newly arisen structure. The gist of what I have tried to show for the development of complement clauses is stated nicely in a recent paper on actualization processes more generally:

> Actualization proceeds from one environment to another on the basis of similarity relations between environments. The similarity relations may involve broad syntactic generalizations but also superficial similarities to existing patterns […] Because actualization is guided by local and global analogies to existing uses, one determinant of the course of actualization is the locus of reanalysis, as it defines the first uses of an item under change, on which subsequent uses can be modeled. (de Smet 2012: 601)

In keeping with this statement, we have seen firstly how the locus of reanalysis determines the *outcome* of actualization in complementation: Despite the fact that lexical diffusion can be quite sweeping and/or involve language-specific idiosyncrasies, it is still often indicative of its very beginnings – so much so that we obtained a quite remarkable cluster of different diffusion patterns and a corresponding match with major classes of diachronic sources of complementation. As far as the *course* of actualization is concerned, a typological study cannot contribute any insights directly, but recourse to language-specific diachronic studies helped to set our cross-linguistic similarity metric of CTP classes into a more revealing light. These studies demonstrated that the similarity relations that drive the diffusion processes of complement clauses are, indeed, to be sought in local and global analogies to existing patterns, just as in de Smet's statement above. Local analogies may be syntactic or semantic in nature. The syntactic ones we explored related primarily to patterns of argument sharing (e.g. subject or object control), while semantic analogies are the well-known ones uncovered by Givón (1980), Ransom (1986), Haspelmath (1989), Cristofaro (2003) and others. In a more recent paper, Cristofaro (2008) integrates formal and semantic aspects of constructional similarity into a usage-based account of how complementation patterns come to spread to novel, especially also unexpected, contexts (such as the Ancient Greek Infinitive in certain types of quotative contexts).[41] Finally, the more 'global' analogies in de Smet's statement may be interpreted as what I called 'systemic' similarities earlier, i.e. parallels in the distribution or usage of functionally related constructions. I referred to Los' (2005) proposals as to how the spread of the *to*-Infinitive in Old and Middle English 'copied' the behaviour of the finite Subjunctive clause, and de Smet (2012) himself provides another nice example from the domain of complementation in English: He discusses extraposed Gerunds (e.g. *It was scary watching it*), which had been around since Early Modern English times, but experienced a dramatic increase in frequency in the early 20th century with a small class of emotive-evaluative predicates (*difficult, great, tough, strange, good*). De

---

[41] I will not go into the specifics of Cristofaro's argumentation here, but the theoretical gist of the account is the following: Language users "abstract particular aspects of the meaning of complex structures as a whole, and associate these aspects of meaning with individual components of complex structures. […] These processes of abstraction and association sanction the reuse of individual components of complex structures in other combinations involving the same or similar meanings." (Cristofaro 2008: 596)

Smet argues that both the initial emergence and the late success of extraposed Gerunds are at least in part explicable by the behaviour of functionally related constructions in the system. Specifically, the emergence of the Gerund in the extraposition construction was most likely influenced by the regular postverbal position of infinitival and *that*-complements (cf. Chapter 5), and the increased usage in the specialized evaluative niche is argued to be due to the rise of formally and semantically similar postverbal *ing*-Participial constructions (e.g. *I feel awful doing that*). The latter had begun to be used with emotive and evaluative predicates in the late 19th century, and the close temporal succession of extraposed evaluative Gerunds is, according to de Smet, unlikely to be a purely accidental pattern. Instead, he argues the specific path of actualization of the extraposed Gerund to be "guided by" or "modelled after" (de Smet 2012: 624) specific usages of the Participial construction. If this is correct, it provides another piece of evidence for the earlier assumption that analogy can be based on concrete and fairly abstract similarity judgements alike.

7.3 Outlook: Organizational differences in complementation systems

In this section, we will wrap up the chapter by taking a brief look at the ways in which individual complementation patterns combine into complementation systems of different kinds. Given that the present sample covers only the major representatives of complementation patterns of each language (cf. §3.2), no system is captured in its entirety and can hence not be analysed in regard to its full ecological organization (for such detailed studies, cf. the contributions to Dixon and Aikhenvald 2006). However, the sample allows for some more coarse-grained observations on certain fundamental typological differences in the organization of complementation systems, and these will be the topic of the present section.

The core issue is, plainly speaking, how 'well-developed' complementation is as a grammatical pattern, i.e. how extensively languages make use of biclausal grammatical constructions for which one could argue that one of the clauses elaborates an argument slot in the other clause. The descriptions of complementation in individual languages soon make clear that there are radical differences here. Contrast, for example, the following kinds of statement:

(245) a. "Sentential complementation is not often used in Mekens. The language resorts to other strategies to encode relationships that are frequently associated with sentential complements in languages like English or Portuguese." (Galucio 2001: 206)

"Complementation is not very common in Hup in the first place" (P. Epps, p.c.)

b. In Karo Batak, "the occurrence of a clausal constituent expounding a nuclear grammatical role of another clause is extremely common." (Woollams 1996: 299)

These impressionistic differences may play out in different ways, which I am going to list first before each of them is briefly turned to individually:

i. Languages differ in the very number of biclausal constructions that may be counted as 'complements', from a single pattern recorded in the present sample "up to half-a-dozen or more" (Dixon 1995: 214) in some languages.

ii. Individual complementation constructions may vary between highly restricted and extremely wide CTP distributions.

iii. Complementation may be "common", in the sense of the above statements, on the level of individual tokens in language use (i.e. a frequently occurring pattern in texts), on the level of types (i.e. the number of distinct complement-taking predicates), or both.

iv. Upon a syntactic delimitation of the phenomenon, a distinction can be made between complement clauses and complementation strategies. Consequently, "some languages work entirely in terms of complement clauses, others only through complementation strategies, while a further set combine the two." (Dixon 2006a: 43)

As for (i), the languages for which only one complementation pattern could be recorded under the present criteria of selection include Kayardild, Kewa, Kiowa, Kwazá, Lavukaleve, Mangarayi, Mandarin Chinese, Mekens, Chalcatongo Mixtec, Somali, Tepehua, Jamul Tiipay, Vietnamese, Warao, Waremwori, Yimas, Yuracaré and Chimalapa Zoque. Additionally, in Kana, Karo Batak, Ainu and Skou, I have conflated several subtypes of complement under a single, overarching complementation pattern since this was homogeneous enough to be covered by the variables in the catalogue. In these languages, then, the core of the complementation system is described with reference to one dominant structural pattern. This is, for instance, a finite subordinate clause in Karo Batak, Kana, Kayardild, Mangarayi, Somali and Zoque; it is a clausal nominalization (at least diachronically) in Ainu, Mixtec, Tepehua, Jamul Tiipay and Yuracaré, a lexical nominalization in Kwazá, Lavukaleve and Yimas, and a (mostly) unmarked sentential structure in Chinese, Mekens, Skou, Vietnamese and Waremwori. In Warao, a deranked construction, involving what is most likely a purposive converbal suffix, "constitutes the nearest equivalent to noun clauses" (Romero-Figeroa 1997: 18), and the only complement-like addition is that of adjunctive direct speech. However, following the policy developed in §2.5, such independent quotative clauses were not considered data points for the present study. This also holds for Kwazá, Lavukaleve and Yimas. Importantly, not all of the above languages can be said to have little-developed complementation systems, as the dominant complementation pattern they have may be extremely versatile and cover a wide range of CTP classes; in other words, criterion (ii) from the above list plays a vital role in assessing the depth of such one-member complementation systems. But before we bring this aspect into the equation, let me briefly mention some languages from the other end of the spectrum, i.e. ones with rather many distinct types of complement (although these may not all be complement clauses in Dixon's strict sense, as we shall see with regard to (iv) below). These include, for example, Choctaw, Hausa, Lezgian, Malayalam, Matsés, Martuthunira, Huallaga Quechua, Santali, Supyire, Tamashek, Tümpisa Shoshone, Tariana and Tzutujil. In such multi-member systems, there is ecological division of labour among the constructions such that each of them is normally adapted to a specific range of constructional meanings (i.e. CTP classes), with little overlap. The general principles on which this ecological organization rests have been described in previous overviews of complementation systems, notably by Noonan (2007: 145ff.) and Dixon (2006a: 23ff.). The consensus between these approaches is that languages typically develop a 'fact' type of complement (called 'indicative' by Noonan) and an

'irrealis' type of complement (called 'subjunctive' by Noonan), and that these may be enriched by an 'activity type', typically for environments of direct perception and similarly predicational (rather than propositional) meanings (this type of complement can be of a 'participial' nature in Noonan's account). These core types of complement can be further differentiated, typically by (additional) non-finite constructions (Noonan 2007: 147). We can see this at work, for example, in Lezgian. This language has a sentential type of complement for indirect speech, propositional attitude and some emotive predicates (cf. (218) again); an Infinitive for either prospective or potential contexts such as SS-want, phasal-P/S ('begin'), emotive-P ('afraid') and causatives; a Masdar mainly for jussive, commissive, certain phasal predicates and various subject clauses (cf. (28)); a 'Substantivized Participle' for perception, declarative knowledge, certain epistemic-S clauses (cf. (40)), and a Converb pattern for DS-want and perception verbs. In other words, Lezgian juxtaposes one finite and four non-finite patterns in complementation, and each of them is used in very specific niches. Noonan's and Dixon's basic distinctions are still visible, but the system is more fine-grained, drawing on various source constructions of the rich subordination system. Huallaga Quechua has exclusively nominalized complements, which are all well-described and sufficiently idiosyncratic to yield four distinct patterns in the sample: a same-subject Infinitive built around an action nominalizer; a more versatile system of nominalization by *-na/-shqa* (comprising both of Dixon's factive and potential type to some degree, cf. (194)); a distinct perception complement expanding an agent nominalizer ((238b) above), and a purposive nominalization for jussive and several other contexts. Tümpisa Shoshone has an unmarked sentential type of complement for all kinds of propositional environments, a quotative complement for indirect speech, an Infinitive and a Participle; each of these constructions has a unique CTP profile with quite idiosyncratic restrictions (cf. the respective records in the database), so again there is a fairly intricate ecological organization at work. More generally, where complementation is a well-developed grammatical pattern to start with, there is reason to believe that the number of distinct constructions is actually much higher than those of the coarse-grained structural types assumed in typological surveys such as Noonan's, Dixon's and my own one. This was alluded to in Chapter 2 (cf. fn. 23 again), based on Cristofaro's (2008) construction-grammatical approach to complementation. Therefore, when we speak of a complementation system with two, three or four members, these are really the most sweeping abstractions over more fine-grained, often lexically-specific, patterns of complementation; assuming these general construction types is justified for comparative purposes, but as was explained in Chapter §3.2, it is questionable that they always have much psychological reality for speakers of the respective language.

The dimension of variation described in (ii) essentially captures how a given biclausal structure, once it has been reanalysed as a complement, penetrates into the complementation system. The co-occurrence matrix that we have been using in this and the previous chapter makes it possible to compare the individual complementation patterns in the sample along this dimension. This was illustrated schematically in Table 19 of Chapter 6, which is repeated in a slightly modified form here for convenience:

Table 23. Co-occurrence matrix of complements and complement-taking environments

	Epis-S	Eval-S	Deon-S	...	SSwant	Know-P	Prop-P	Quot-P	Other-P	CIP_{pat}
Mixtec-*xa=*	1	1	1	...	1	1	1	1	1	16.1
Mrt2	0.1	0.1	0.1	...	1	0	0.5	0.5	0.1	4.0
Toq-*na*	0.1	1	1	...	1	1	1	0.9	1	9.9
...	...	...	...	...	...	...	...	...		...
CIP_{class}	63.9	117.9	71.8	...	90.7	102.7	105.9	101.9	146.0	...

The very last column of the table contains the cumulative productivity index (CIP) for each complementation pattern in the data. Productivity in the present sense refers to the number of CTP *classes* that a given complement can co-occur with; importantly, it makes no claim as to the total number of individual complement-taking predicates, so it is not a measure of type frequency in the classic, corpus-linguistic, sense. Recall also from Chapter 6 that I introduced a class called 'other-P' (cf. the last-but-one column in Table 23) in order to capture that a given pattern may be productive in the P-domain beyond the specific environments defined for the present study. This CTP class has played no role so far but it must, of course, now be included in the CIP calculation for each data point (even if it is only a very rough and impressionistic estimate). If we take the CIP_{pat} as an approximation of the degree of diffusion of each complement, we can see that the CIP_{pat} values in the sample vary widely, between 1.3 and 17.1 (mean = 6.97, *SD* = 3.57). Extremely low values (e.g. $CIP_{pat} \leq 2.0$) imply that some constructions cling very firmly to the source context in which they arose, failing to diffuse in the complementation system. This is characteristic, for instance, of the quotative complements in Choctaw, Mapudungun, Matsés, Martuthunira or Tümpisa Shoshone, of a purposive construction in Santali (which is used as a complementation strategy only for jussive verbs), or of the perception complement in Huallaga Quechua discussed earlier (cf. (238b), presumably arisen through the matrix coding pathway). By contrast, extremely high values are indicative of particularly successful actualization, i.e. a wide propagation of the complement. These are precisely the constructions we identified in the central area 'E' of the *NeighborNet* graph discussed earlier (cf. Material 8 again), and they include the unmarked complements in Karo Batak (optionally enhanced by a purpose marker), Mandarin Chinese and Vietnamese, the (probably purpose-based) finite complement clause in Serbo-Croatian, the relative-based complements in Persian and Georgian, and the nominalizations in Chalcatongo Mixtec (clausal) and Musqueam (clausal/lexical).

The CIP_{pat} figures become more interesting when we relate them to the structure of complementation systems as discussed above. We can now see, for example, that the complements of one-member systems differ dramatically in their productivity. The unmarked sentential structures from Batak, Chinese and Vietnamese just mentioned can virtually cover the entire range of CTP environments; to be sure, one could distinguish certain subpatterns in all three of these languages again, based on the distribution of optional subordinators (cf. the database), but the unmarked superimposition of a sentence-like structure is possible across the board, precisely because the absence of any marker whose history could potentially constrain the application of the pattern. But even where such a marker exists, widespread coverage is perfectly possible, as in Kana (quotative complement) or Mixtec (clausal

nominalization). Other one-member systems are very different because the one unquestionable type of complement they have developed is still restricted in its distribution. This can be seen, for instance, in Warao, Kiowa or Mangarayi. In Warao, as mentioned above, a purposive converb is used to cover same-subject desiderative environments (cf. (58) in Chapter 4), and we do not have explicit evidence of any other genuine complementation contexts; the primary means for linking propositions or states of affairs in this language is that of "parataxis" (Romero-Figeroa 1997: 12). In Mangarayi, "we find a single formal subordinate clause type which is variably understood as adnominal or adsentential modifier", the specific interpretation depending on the "interaction between formal and textual features" (Merlan 1982: 13). This 'generalized subordinate clause' was illustrated in (234) above, where we noted that it can be indistinguishable between a temporal and a complement reading after perception verbs. Crucially, apart from perception (and possibly emotive-P) contexts, all other CTP classes are rendered by alternative means in Mangarayi, such as desiderative affixes, modal particles, causative auxiliaries, adjunctive direct speech, etc. (cf. the database for the details, and recall the comments on Mangarayi made in §6.3.2). Other one-member systems score somewhere in between the two extreme poles.

Constructions that contribute to multi-member systems with a highly diversified internal structure, like Choctaw, Lezgian, Huallaga Quechua, Santali, Supyire, Tariana, Tzutujil, Tümpisa Shoshone, etc. are characterized, quite expectedly, by relative low CIP values. As discussed above, the division of labour expounded by these systems often entails that the individual patterns are specialized for their respective ecological niches, so that no across-the-board extension takes place. Consequently, their CIP values often range within reasonable limits, e.g. between 2.0 and 4.5 in Tümpisa Shoshone, between 2.9 and 5.7 in Tariana, or between 4.1 and 6.0 in Supyire. Sometimes, there is a still a dominant pattern with a wider distribution (e.g. Choctaw 9.7, Tzutujil 9.8, Quechua 9.8), but this is still quite far from the extremely productive constructions we encountered above (e.g. with CIPs above 14).

In regard to criterion (iii), i.e. the specific type and token frequency of complementation in each language, our CIP calculations cannot offer any direct insights, as they are class- rather than item-specific measurements. It is clear from inspecting the reference grammars, however, that languages are fundamentally different here. In some, the number of predicates that can enter into complementation relations is very large (e.g. Gulf Arabic, Karo Batak, German, Hausa, to name but a few), while there are explicit statements in others that such lexicalization patterns are more restricted. For Menya and Kewa, for example, we know that the CTPs entering complementation "are of a small set" (Franklin 1971: 121 on Kewa). On a construction-specific level, the extreme case of type restrictions is one where the complement is lexically specific, i.e. applicable to a single CTP. In Matsés, for instance, the predicate *bun* 'want' is the only predicate to which a certain complement clause can apply (cf. (15b) in Chapter 2); and more generally, "there are relatively few complement-taking verbs" in the language as a whole. "Part of this has to do with the fact that there are highly polysemous verbs that cover multiple complementation-type notions." (Fleck 2006: 243) This situation is similar, and even more extreme, in

Ungarijin, where we find an invariant intransitive verb root *ma* which can combine with a clause in a complement-like relationship and, depending on the context, yield a variety of different meanings (e.g. causative, jussive, desiderative, quotative):

(246) Ungarinjin (Australian, Wororan: Australia; Rumsey 1982: 162)

  [*Wuḷan wurumiyaŋga*]   *budmaṟaŋarugu.*
  word 2:CLASS.3PL.FUT.know 3PL.√ma.PST.1PL:IRR.BEN
  '"They will know this word", they did with regard to us', or
  'They wanted us to know that word.', or 'They made us know that word.'

In other words, here, too, there is only a single CTP, though it can express a variety of constructional meanings. (The opposite of this is typical of some quotative constructions, which can attract a number of different verbs of saying ('say', 'speak', 'shout', 'ask') but only a single one of our constructional meanings, i.e. the utterance class.) Exhaustive listings and discussion of all individual CTPs can be found in some of the reference materials (such as Lichtenberk's (2008) study of To'aba'ita, which carefully walks the reader through virtually all CTPs), and again in the contributions to Dixon and Aikhenvald (2006).

We can now proceed to the final dimension of variation listed above, namely differences in the syntacticization of complement relations. At various points in this dissertation, we encountered the distinction between complement clauses (CC) and complementation strategies (CS), set out most clearly in Dixon (2006a). It needs to be recalled that I have drawn the boundaries between these two terms differently from Dixon's original proposal, in several respects: (i) Dixon imposes a constraint that a complement clause must replace a nominal argument in the main clause; therefore, clauses required after intransitive verbs are never CCs. By contrast, I have counted them as such if a morphologically intransitive verb requires the presence of the subordinate clause, even though it never requires the presence of an object NP. (ii) Dixon makes the distinction contingent on the kinds of CTPs that occur in a given constructions – if a certain set of core CTPs happens not to be covered, the construction is considered a complementation strategy (cf. §2.5 again). As explained earlier, I simply lifted this constraint, as it did not seem motivated enough and could not be operationalized rigorously. (iii) If there was evidence that at least one CTP takes a complementation pattern as its syntactic argument, I counted it as a complement clause, even if it is a complementation strategy for the remaining ones. (I do not know how Dixon would proceed in such cases.) An exception is that of direct speech, which Dixon generally excludes from the domain of complementation, but which I took on board as a complementation strategy if it had argument status and if there was no indirect-speech complement.

With these modifications in mind, we can now examine how many of the sample languages exhibit genuine complement clauses (in my understanding) and which ones need to be characterized exclusively, or predominantly, in terms of complementation strategies. Since I adopted a more liberal policy, the number of languages for which *all* relevant constructions in my data are best considered complementation strategies is extremely small. One such language is Kiowa, already mentioned above, in which a construction for indirect speech and thought is juxtaposed to the main clause and no

embedding could be discerned; this was discussed in §6.4.1. Another case in point is Yimas, which uses independent discourse units (parataxis) for a number of the propositional complement relations in other languages, so that a nominalized construction is the only genuine complementation pattern in the data. Crucially, though, these nominalizations "are clearly noun phrases" (Foley 1991: 394). Having all internal and external properties of NPs, they fail Dixon's 'clausal' criterion and hence must also be considered complementation strategies. On this account, then, Yimas does not have complement clauses. Incidentally, a similar statement ("There are no complement clauses") has been made by Terrill (2003: 351) on Lavukaleve, but here the nominalization clearly retains some verbal syntactic properties and thus needs to be considered a complement clause on Dixon's and my account. But there is another language from the Papuan area which may have to be assumed to lack complement clauses, viz. Menya. This language employs a clause-chaining strategy for perception verbs, a left-adjoined subject clause pattern (cf. §5.2 again), and a direct-speech complement of a ditransitive verb 'say, tell'. As was stated above, I counted the latter preliminarily as a CS, but in fact, its argument status may justify treating it even as a genuine CC. (Dixon would discard it altogether because it is direct speech.) Overall, then, Menya may turn out to have CCs, after all. On this view, Menya would group with languages in the sample in which CSs are generally more prominent than genuine CCs. One of them is the Pama-Nyungan language Martuthunira, in which complements "are usually [...] relative clauses" or purpose clauses (Dench 1995: 255). As such, they are typically adjuncts to either an argument of the matrix clause or the matrix clause as a whole. Consequently, Dench argues that these subordinate clauses "are adjoined rather than embedded" (ibid.: 240) and thus constitute complementation strategies rather than complement clauses in Dixon's sense. Interestingly, there appears to be one exception to this adjunctive organization of the complementation system, relating to the utterance verb *wangka* 'say'. This verb can be used intransitively, in the sense of 'making a noise', but also transitively or even in a double object construction, with both the addressee and the theme argument marked for accusative case (lit. 'I told him-ACC the story-ACC'). A clause as the theme argument of *wangka* can be found in (247) below:

(247) Martuthunira (Australian, Pama-Nyungan: Australia; Dench 1995: 223)

Yartapalyu	*wangka-nguru*	[*parna-ngka-rru*	*kangku-lha-a*],
others	say-PRS	head-LOC-now	carry-PST-ACC

yartapalyu	*wangka-nguru*	[*warryayi-lalha-a*].	
others	say-PRS	drag-PST-ACC	

'Some say they carried it on their heads, others say they dragged it.'

This is what Dench (ibid.) calls an 'indirect speech complement". It works differently from the above complementation strategies because it is not a 'pivotal' construction; in contexts such as (247) above, there is no chance that the accusative marking on the subordinate verb signals coreferentiality with an accusative object in the matrix clause (because there is none in the present example). Therefore, the most likely analysis, also endorsed in a recent cross-linguistic study of quotative constructions (Spronck

2012: 93), is that the complement functions directly as the object of 'say' and hence constitutes a genuine complement clause. This organizational pattern is special from a cross-linguistic point of view since it is more typical, as we saw above, for the quotative construction to be adjunctive in nature while other complements are genuinely embedded. In Martuthunira, the situation is reversed. Another language in which adjunction prevails over embedding is Santali. The language has a reduced type of complement which can function as the object argument of certain predicates; in addition, however, there are three complementation patterns which, at least in the P-domain, are predominantly adjoined to a syntactically saturated matrix clause and hence are CS on most occasions. This is because of extensive prolepsis, both in a relative-like structure ('I saw her [die]') and in an extended quotative construction (cf. (219) above: 'You know (about) me that I'm the big one'), or due to adjunction of a purpose clause in jussive environments. In other languages, several kinds of CS conspire to outweigh genuine complement clauses. For example, Mapudungun, Motuna and To'aba'ita each have one CC, a strong nominalization as a CS, and a quotative construction that needs be considered a CS by our criteria: it is direct-speech 'complement' (not adjunct) in Mapudungun and Motuna, and a syntactically adjoined clause for indirect speech and all kinds of other CTPs (jussive, causative, perception, cf. (5) in Chapter 2) in To'aba'ita. At several points in the dissertation, we noted that embedded complement clauses can develop from adjoined clauses by virtue of dropping a coreferential antecedent in the matrix clause. This process appears to be taking place in Supyire. The language has an Indicative and a Subjunctive type of complement, and as was discussed in Chapter 5 (cf. (103a)), both of these "arose through parataxis" (i.e. in right-adjoined position) and have "since developed a number of characteristics which show that they have to varying degrees become more closely integrated into the syntax of the main clause" (Carlson 1994: 465). Specifically, where a cataphor in the matrix clause is missing, these clauses may be interpreted as functioning directly as the object or subject argument of the respective CTP. In view of such performance alternations, "Supyire complement clauses are at an interesting stage between parataxis and full-scale embedding" (ibid.: 460).[42] More generally, we can here see the difficulty again, also discussed in Chapter 5, of deciding on the CC/CS status in languages with pro-drop possibilities for subjects and/or objects. For the present section, I have taken the absence of any overt pronominal antecedent as indicative of embedding (and hence CC status), but this decision can, of course, be questioned. If it is, then the number of complementation strategies would certainly go up, and a language like Epena Pedee, for example, would come out as lacking complement clauses. Epena Pedee is an OV language and object complements are unmarked postverbal constructions:

---

[42] Note that, in addition to the two above-mentioned complements, Supyire also has a further, proleptic (and hence adjoined) type of CS for perception and causative environments, as well as a strong nominalization as a CS. Overall, then, the system is on its way to becoming a mixed one hosting both CCs and CSs.

(248) Epena Pedee (Chocó: Colombia; Harms 1994: 173)

Nepɨrɨ-pa-či-dá [*rey táu pʰãríu bɨ*].
tell-HAB-PST-PL king eye dark be
'They told him the king was blind.'

Since Epena Pedee as a pro-drop language, it is unclear if we need to assume an implicit cataphor of the object clause in the matrix; this was also discussed in relation to a subject-clause construction in §5.5 (cf. (142)). For deciding on the CC/CS status of these complements, then, the 'cataphor issue' is of great significance.[43]

Overall, the distinction between CCs and CSs is, I think, a useful one for exploring the typological differences in complementation systems. However, in its current conceptualization, it is a multidimensional distinction that rests on debatable decisions along each dimension: Where exactly is the boundary between arguments and adjuncts? When exactly does a nominalized structure count as an NP? Do we assume invisible placeholders as viable antecedents of complement clauses? What is the status of direct speech in a typology of complementation? It is the answer to each of these questions that influences the assessment of individual languages as lacking complement clauses in the narrow sense. And since a given construction can be embedded in one syntactic function but adjoined in another, the issue is probably best addressed at the level of such functions. (In §6.4, I provided such an analysis for S- and A-environments.) Therefore, the question of whether a given language works entirely in terms of complement clauses or complementation strategies is a challenging one to answer from a theoretical point of view, let alone in comparative perspective.

Bringing together all considerations from (i) to (iv) above, and the specific problems engendered by each of them, it is a far from trivial task to operationalize how well-developed complementation is as a grammatical category in the sample languages. Dixon himself (1995: 183) states that "from general reading, and from talking to descriptive linguists, I have the impression that complement-clause constructions are common among the languages of Europe, Oceania and Africa but rare in those of Australia and South America." Drawing on the discussion of the four criteria above, this impression appears to be generally confirmed. The languages for which prose descriptions suggest little-developed complementation systems tend to be found in either South America or Australia-Papua New Guinea, and the languages with a particular propensity for complementation strategies (in my somewhat different understanding of the term) have the same bias (except for Supyire and Santali). The sample languages from Papua New Guinea are mostly characterized by having complementation patterns with low to medium CIPs. In Imonda, for example, there are two constructions with fairly specialized niches (CIPs = 3.3 and 3.4, respectively) and nothing comparable to a widely diffused complement clause of more familiar languages. Similar situations arise in Menya, Lavukaleve, Yimas, Kewa and Motuna, all

---

[43] The issue is similar in Persian, where *ke*-complements do not appear in the canonical preverbal position unless they are nominalized by a head noun or demonstrative (cf. Chapter 5). Therefore, one may argue that the standard postverbal construction is not a proper complement clause, by appears in apposition to pro-dropped cataphor in the matrix. However, in the literature, it is normally treated as a complement clause ("Postverbal CPs which appear by themselves [i.e. without a cataphor, KSB] are complements." (Aghaei 2006: ix)).

of which have been mentioned in connection with restrictions on complementation. In Korafe, Amele, Abun, Skou and Warembori, complement clauses of the more familiar type are attested and also achieve a wider CTP distribution (e.g. CIP_{pat} = 8.7 in Amele, 9.0 in Skou or 9.9 in Korafe). More generally, however, it is telling that, for example, Foley's survey of the organization of Papuan languages is completely reticent on complementation in its discussion of clause combining (cf. Foley 1986: 175–205), indicating that complementation is not a prominent category in these languages, by whichever one of the criteria in (i) to (iv). For Australia, Dixon (2006c: 263) is even more determined, claiming that "a complement clause construction is found in few (if any) Australian languages." It is certainly true that Australian languages tend to have relatively weak complementation systems. We noted at several points that the so-called 'generalized subordinate clauses' can be used to evoke a complement reading (e.g. Mangarayi, Wambaya, Ungarinjin), but it can be hard to identify whether they are adjoined or embedded constructions in such contexts. It seems to me that some of the Australian complements have undergone Hale's (1975) pathway from adjunction to embedding in specific niches (e.g. Wambaya, Gooniyandi, Ungarijin), including the formerly pivotal or 'anchored' complements in Kayardild that have now developed into genuine complements with 'complementizing case' rules (cf. §4.4 again). In Martuthunira, we can still see the source constructions at work, so that the language operates largely (though not exclusively) in terms of adjunction. As in the Papuan languages, the CIPs of individual complementation patterns tend to be rather low in Australia as well (e.g. 3.4 in Mangarayi, 5.0 in Wambaya, 1.9–4.0 in Martuthunira). In view of this situation, Dixon's impression is, of course, not misguided, although a more thorough investigation is needed to substantiate if the majority of Australian languages really lacks complement clauses in the narrow sense; again, this depends on the specific definition of the CC/CS boundary, but at least in terms of argument status, I do see some potential for CCs even in Australian languages.

As for South American languages, a notable pattern is the co-existence of two construction types on the outskirts of complementation, namely lexical nominalizations on the one hand and direct speech on the other. Depending on how far each of these moves into the direction of typical complements (e.g. by the embedding and diffusion of direct speech as a complement, or by verbalization of the nominal constructions), we sometimes get genuine complement clauses even in those languages (e.g. Mapudungun, Awa Pit, Barasano, Hup, Jarawara, Kwazá, Matsés and others). The more verbal types of nominalization can even come to achieve a relatively high CIP (e.g. Mapudungun 11.9, Jarawara 12.4, Kwazá 10.7). Overall, however, we find lots of low-profile, or highly restricted, complements in this macro area, and very few across-the-board complements of the Eurasian type. Given the prevalence of nominalization and direct quotation, a thorough assessment of the situation is again reliant on where one wishes to draw the CC/CS boundary. Recall from an earlier discussion that Dixon himself considers indirect (!) speech in Jarawara as a CS because of its limited CTP distribution, despite the fact that is a proper argument of a matrix predicate. In my operationalization, by contrast, there are a lot more complement clauses in South American languages, and virtually no systems that work *entirely* in

terms of complementation strategies (Epene Pedee is debatable, as we saw above). Instead, many South American languages show a fairly even opposition between CCs and CSs (e.g. Hup, Matsés, Mosetén, Sanuma or Tariana). Note that the situation in North America (including most of the Mesoamerican languages in the sample) is different. Here, the prevalence of clausal nominalization, and clearer instances of subordinate clauses in argument positions more generally, lead to more systems with CCs only, at least in my understanding of the term (e.g. Lakota, Wappo, Slave, Mixtec, Yuchi, etc.). And as we saw in the *NeighborNet* exploration of the data, some of the relevant constructions (e.g. from Mixtec, Chumash and Musqueam) are among the most widely diffused patterns in the sample (with CIPs between 14 and 16.1), thus having achieved the status of a versatile complement (and subordinate) clause that appears to be rather rare in South American languages.

I shall confine myself to these observations here. Obviously, "systematic study […] of the genetic and areal distribution of the various kinds of clause linking" is a pressing issue for future research, as Dixon (1995: 183) points out. However, this was not envisaged by the present section (nor the dissertation as a whole, since different kinds of sampling and selection techniques would have been necessary to tackle these issues). I have rather tried to sketch the various dimensions in which a complementation system may be said to be poorly developed, and to relate previous statements on the geographical distribution of such systems to the languages in the present sample. In view of the conceptual difficulties involved in approaching and measuring the organization of complementation systems, this could be no more than an 'outlook', as suggested by the title of the section.

Envoi

Noonan (2007: 150) concludes his seminal survey article on complementation by saying that "there are a large number of works which deal with aspects of complementation in individual languages, but few that provide an overview of complementation, either cross-linguistically or in a particular language." The present dissertation was meant to make a contribution to improving this situation for the cross-linguistic part of Noonan's statement. Specifically, my goal was to take the two major typological survey works on complementation, i.e. Noonan (1985|2007) and Dixon (2006a), in order to investigate some of their central conceptual distinctions, distributional hypotheses and hitherto understudied aspects against a thoroughly empirical background. In doing so, I have exploited data from over 100 languages, which, despite the qualitative drawbacks pointed out in Chapter 3, provided a wealth of interesting material to further enhance our understanding of the grammar of complementation. The dissertation was divided into four interconnected case studies on the topic, all of which focused, in one way or another, on the organization of complementation patterns into specific grammatical (sub)systems. One of the most important aspects in this connection was the distribution of complement clauses over different argument-structural functions, and the morphosyntactic and syntagmatic constraints that tend to become operative across the world's languages when complements are extended to functionally dispreferred environments, notably the 'subject' function of certain matrix predicates. The relevant findings have already been summarized and connected to each other in the previous chapters, and I do not wish to drag out the dissertation by reiterating the same points here. Instead, I will conclude with a few 'meta'-comments on the research project as such.

One important aspect of the present research that needs to be pointed out is that it offers but one specific *interpretation* of the cross-linguistic data on complementation. It is heavily influenced by the specific sampling and selection procedure, by the materials consulted, and the coding decisions made all along. I have tried to make each of them transparent to the reader, but we have seen that the goal of 'measuring' (dis)similarities in the grammatical organization of complementation requires one to translate the cross-linguistic data into discrete, and sometimes even numerical, levels of operationalization, which are bound to be debatable. My intention in doing so was to explore how measures like the degree of desententialization ('CID') or the degree of productivity ('CIP') of complement clauses and complement-taking environments can be used to detect structural, distributional and diachronic regularities in a large set of sample languages. In this way, I have tried to set the typological study of complementation on a broader and more differentiated empirical basis than previous

overviews, and it remains to be seen whether the results can stand the test of more detailed future enquiries along similar lines.

In making a first step towards such a broader empirical picture, the study has probably raised more questions and issues than it has been able to answer in a straightforward way. In many places, the dissertation has actually been a problematization of doing typological research on complement clauses. For example, I have emphasized the general usefulness of Dixon's distinction between complement clauses and complementation strategies, but it has also become apparent that it is itself dependent on many debatable and presently unresolved issues, such as the argument-adjunct distinction, the status of direct speech in a typology of clause combinations, the gradient nature of the NP–clause distinction, the status of pro-dropped placeholders etc. It is only on very consistent and cross-linguistically commensurable application of these criteria that the true extent of similarities and differences in complementation systems becomes visible, and the present study could not offer a fully satisfactory solution to these issues either. In parallel to settling these theoretical questions, the most pressing issue that needs to be tackled is a refinement of the database on which the typological study of complementation draws. As was discussed in Chapter 3, my own data presently reflect the information gathered from grammatical descriptions, enhanced on occasion by informants' and experts' opinions on some of the variables in question. This was a rich foundation for the first exploratory purposes of the present research, but the future study of complementation from a typological perspective would clearly benefit from a larger, coordinated database to which experts on individual languages can contribute directly, based on the corpora and elicited material available to them (similarly, for example, to the *ValPaL* database (Hartmann et al. 2013)). The present investigation has shown that some aspects of complementation systems tend to lack systematic coverage in grammatical materials and that, at present, we cannot be entirely sure whether this reflects the genuine absence of certain phenomena from the languages in question or simply a gap in the materials. Above all, this applies to complement clauses in subject function: While there is reason to believe that S- or A-clauses are more restricted across languages than P-complementation, as laid out in Chapter 6, they may still be more widely attested than the current study suggests. Clearly, what we need is more precise information on the syntactic functions in which complement clauses can occur in a given language (in the spirit of the contributions to Dixon and Aikhenvald 2006), what the relevant coding frames are (canonical/non-canonical argument structures), and by which criteria the complement could be said to be a syntactic argument to begin with. And in all of this, we need to achieve broader genealogical coverage, so as to iron out the skewing of the present sample, as well as a better understanding of the areal dynamics that lead to convergence in complementation. In this light, the dissertation and its underlying data must be seen as the beginning, rather than the finished product, of a typology of organizational patterns in complementation systems.

Appendix

Material 1: Sample languages (by genetic affiliation in alphabetical order) and informants

Afro-Asiatic
    Berber: **Tamashek** (Jeff Heath)
    Chadic: **Hausa** (Mahamane Abdoulaye)
    Cushitic: **Somali**
    Omotic: **Wolaytta**
    Semitic: **Gulf Arabic** (Clive Holes)
Altaic
    Turkic: **Turkish**
    Tungusic: **Evenki** (Igor Nedjalkov)
Arauan: **Jarawara**
Araucanian: **Mapudungun** (Fernando Zúñiga)
Arawakan: **Tariana**
Australian
    Bunuban: **Gooniyandi** (Bill McGregor)
    Mangarrayi: **Mangarayi**
    Pama-Nyungan: **Martuthunira**
    Tangkic: **Kayardild**
    West Barkly: **Wambaya** (Rachel Nordlinger)
    Wororan: **Ungarinjin**
Austro-Asiatic
    Mon-Khmer/Aslian: **Semelai**
    Mon-Khmer/Viet-Muong: **Vietnamese** (Ngoc Thuy Duong Dao)
    Munda: **Santali** (Lukas Neukom)
Austronesian
    Central Malayo-Polynesian: (Fehan) **Tetun**
    Eastern Malayo-Polynesian/Oceanic: **To'aba'ita**
    Eastern Malayo-Polynesian/ South Halmahera-West New Guinea: **Taba**
    Western Malayo-Polynesian/Sulawesi: **Tukang Besi** (Mark Donohue)
    Western Malayo-Polynesian/Borneo: **Begak Ida'an**
    Western Malayo-Polynesian/Sundic: **Karo Batak** (Geoff Woollams)
    Western Malayo-Polynesian/Sama-Bajaw: **Yakan**
Barbacoan: **Awa Pit**
Border: **Imonda**
Chapacura-Wanham: **Wari'**
Chibchan: **Rama**
Chocó: **Epena Pedee**
Chumash: **Barbareño Chumash**
Dravidian: **Malayalam**
East Bougainville: **Motuna**
Eskimo-Aleut: **West Greenlandic** (Michael Fortescue)
Hmong-Mien: **Hmong Njua** (Bettina Harriehausen-Mühlbauer)
Hokan/Yuman: **Jamul Tiipay**
Indo-European
    Germanic: **German**
    Iranian: **Persian** (Peter Öhl; various native speakers)
    Slavic: **Serbo-Croatian** (Wayles Browne, Margita Soldo)
Kartvelian: **Georgian** (Merab Geguchadze)
Khoisan: **(Modern) Khwe** (Christa Kilian-Hatz)
Kiowa-Tanoan: **Kiowa**
Lower Mamberamo: **Warembori**
Lower Sepik-Ramu/Lower Sepik: **Yimas**
Makú/Vaupés-Japurá: **Hup** (Pattie Epps)
Mayan: **Tzutujil**
Mixe-Zoque: (Chimalapa) **Zoque** (Heidi Johnson, Terje Faarlund, Roberto Zavala Maldonado)
Mosetenan: **Mosetén**
Muskogean: **Choctaw**
Na-Dene/Athapaskan: **Slave** (Keren Rice)
Nakh-Daghestanian/Lezgic: **Lezgian**
Niger-Congo
    Northern Atlantic: **Noon**
    Benue-Congo/Cross-River: **Kana**
    Benue-Congo/Bantoid: **Nkore-Kiga**
    Dogon: **Jamsay** (Jeff Heath)
    Kwa: **Fongbe** (Clair Lefebvre)
    Adamawa-Ubangian: **Mayogo** (Kenneth Sawka)
    Gur: **Supyire**
Nilo-Saharan
    Kadugli: **Krongo**
    Moru-Ma'di: **Ma'di**
    Nilotic: **Lango**
    Songhay: **Koyra Chiini**
Oto-Manguean/Mixtecan: **Chalcatongo Mixtec** (Monica Macaulay)
Panoan: **Matsés**
Peba-Yaguan: **Yagua**
Quechuan: **Huallaga Quechua**
Salish/Central Salish: **Musqueam**
Sino-Tibetan
    Tibeto-Burman/Bodic: **Dolakha Newar** (Carol Genetti)
    Tibeto-Burman/Burmeso-Lolo: **Burmese**
    Sinitic/Chinese: **Mandarin Chinese** (Martin Schäfer)
Siouan: **Lakota**
Skou/Western Skou: **Skou**
Solomons East Papuan: **Lavukaleve** (Angela Terrill)
Tai-Kadai/Kam-Tai: **Lao**
Tarascan: **Purépecha**
Totonacan: **Tepehua** (Susan Kung)
Trans-New Guinea
    Angan: **Menya**
    Binanderean: **Korafe**
    Engan: **Kewa**
    Madang: **Amele** (John Roberts)
Tucanoan: **Barasano**
Tupian/Tupí-Guaraní: **Mekens**
Uralic: **Hungarian**
Uto-Aztecan/Numic: **Tümpisa Shoshone**
Wappo-Yukian: **Wappo**
West Papuan/North-Central Bird's Head: **Abun**
Yanomam: **Sanuma**
Yukaghir: **Kolyma Yukaghir**
Isolates
    Ainu (Anna Bugaeva)
    Basque
    Japanese (Kyoko Maezono, Toshio Ohori)
    Korean
    Kwazá
    Urarina (Knut Olawsky)
    Trumai
    Warao
    Yuchi
    Yuracaré (Rijk van Gijn)
Pidgins and Creoles: **Ndyuka**

Material 2: Major printed sources of information on the sample languages

The abbreviation "C" indicates comparative sources (e.g. on other varieties of the sample language or on other members of the same stock; comparative literature of a more general kind is not typically included here).

Abun	Berry and Berry 1999, Berry 1995; **C:** Dol 1999, Hayami-Allen 2001, Holton 2003
Ainu	Refsing 1986, Tamura 2000, Bugaeva 2008a, 2008b, 2012; Shibatani 1990
Amele	Roberts 1987, 2001
Arabic (Gulf)	Holes 1990; **C**: Brockelmann 1980, Kremers 2003, Persson 2002, Reckendorf 1898
Awa Pit	Curnow 1997, 2002
Barasano	Jones and Jones 1991
Basque	Hualde and Ortiz de Urbina 2003, Saltarelli 1988, Trask 1997
Batak (Karo)	Woollams 1996
Begak Ida'an	Goudswaard 2005
Burmese	Soe 1999
Choctaw	Broadwell 2006, Davies 1981, Nicklas 1974
Chumash (Barbareño)	Wash 1999, 2001, Ono 1996
Epena Pedee	Harms 1994
Evenki	Nedjalkov 1997, Grenoble 2012
Fongbe	Lefebvre and Brousseau 2002, Höftmann 1993
Georgian	Hewitt 1995, Hewitt 1987, Amiridze 2006, Aronson 1990, Harris 1981, Skopeteas et al. 2007, Vogt 1971
German	Eisenberg 2006, Fabricius-Hansen 2005, Zifonun 1997, Inaba 2007, Mollica 2010, Paranhos Zitterbart 2002
Gooniyandi	McGregor 1990, 1988
Greenlandic (West)	Fortescue 1984; **C**: Jacobson 1995
Hausa	Jaggar 2001, Newman 2000, Abdoulaye 2008, 2009, 2011, Bagari 1971
Hmong Njua	Harriehausen 1990, Taweesak 1984; **C**: Jarkey 2006
Hungarian	Kenesei et al. 1998, Kenesei 1994, Keresztes 1993, Kiss 2002, MacWhinney and Pléh 1988, Rebrus and Babarczy (n.d.), Szabolcsi 2009, Szent-Iványi 1995, Benkö 1993, De Groot 1995
Hup	Epps 2008
Imonda	Seiler 1985
Jamsay	Heath 2008
Japanese	Hinds 1986, Kuno 1972, 1973, Kaiser et al. 2013, Wrona 2005, 2007, 2008, Coulmas 1986, Maynard 1986, Yap and Matthews 2008
Jarawara	Dixon 2004, 2006b
Kana	Ikoro 1996
Kayardild	Evans 1995
Kewa	Franklin 1971
Khwe (Modern)	Kilian-Hatz 2008, Kilian-Hatz 1999
Kiowa	Watkins 1984
Korafe	Farr 1999
Korean	Sohn 1994, 2009, Chung 1995, Gamerschlag 2007, Lee 1994, Park 1993, 1995, Yang 1994, Yap and Matthews 2008
Koyra Chiini	Heath 1999
Krongo	Reh 1985
Kwazá	Van der Voort 2004
Lakota	Buechel 1939, Van Valin 1977
Lango	Noonan 1992, 2007
Lao	Enfield 2007
Lavukaleve	Terrill 2003
Lezgian	Haspelmath 1993
Ma'di	Blackings and Fabb 2003, Andersen 1998
Malayalam	Asher and Kumari 1997, Andronov 1996, Jayaseelan 1999, 2004, Krishnamurti 2003, Philipose 1984, Shibatani 1999, Sridhar 1976, Sundaresan and McFadden 2009
Mandarin Chinese	Li and Thompson 1981, Yue 2003, Chappell 2008, Dryer 2003, Huang 2003, Wang 2006
Mangarayi	Merlan 1982
Mapudungun	Smeets 1989, 2008, Zúñiga 2006
Martuthunira	Dench 1995
Matsés	Fleck 2003, 2006
Mayogo	Sawka 2001
Mekens	Galucio 2001, 2002
Menya	Whitehead 2004
Mixtec (Chalcatongo)	Macaulay 1996, 2005; **C**: Hollenbach 1995, Johnson 1988
Mosetén	Sakel 2004
Motuna	Onishi 1994, 2000, 2004
Musqueam	Suttles 2004, Kroeber 1999, Thompson 2012; C: Gerdts and Hukari 2007, Kiyosawa 2006
Ndyuka	Huttar and Huttar 1994

Nkore-Kiga	Taylor 1977, 1985
Noon	Soukka 2000
Newar (Dolakha)	Genetti 2006, 2007, Genetti et al. 2008
Persian	Mahootian 1997, Aghaei 2006, Dehghani (n.d.), Lambton 1974, Mace 2003, Lotfi 2008, Öhl and Lotfi 2007, Windfuhr 1979
Purépecha	Chamereau 2000, Foster 1969
Quechua (Huallaga)	Weber 1989, 1994; **C**: Muysken 1994, Adelaar and Muysken 2004, Miller 1989
Rama	Grinevald (n.d.)
Santali	Neukom 2001
Sanuma	Borgman 1990
Semelai	Kruspe 2004
Serbo-Croatian	Kordić 1997, Belić 2005, Browne 1987, 1993, Browne and Alt 2004, Joseph 1983, Kim 2010, Kordić 1999
Skou	Donohue 2004
Slave	Rice 1989, 1986
Somali	Saeed 1999, Berchem 1991, Kirk 1905
Supyire	Carlson 1994
Taba	Bowden 1997, 2001
Tamashek	Heath 2005
Tariana	Aikhenvald 1999, 2003, 2006
Tepehua	Kung 2007, Watters 1985, 1988
Tetun (Fehan)	Van Klinken 1999
Tiipay (Jamul)	Miller 2001
To'aba'ita	Lichtenberk 1991, 2008
Trumai	Guirardello 1999
Tümpisa Shoshone	Dayley 1989
Tukang Besi	Donohue 1999b, Klamer 2000
Turkish	Kornfilt 1997, Göksel and Kerslake 2005, Lewis 1967, Underhilll 1976; Csató 2010, Erdal 2004, Ersen-Rasch 2004, Hennesy and Givón 2001, Johanson 2011, Słodowicz 2007
Tzutujil	Dayley 1985
Ungarinjin	Rumsey 1982
Urarina	Olawsky 2002, 2006
Vietnamese	Nguyên, D. 1988, 1996, 1997; Nguyên, Th. 1979, Thompson 1965; Bisang 1992, Clark 1996
Wambaya	Nordlinger 1998, 2002, 2006
Wappo	Thompson et al. 2006
Warao	Romero-Figeroa 1997
Warembori	Donohue 1999a
Wari'	Everett and Kern 1997
Wolaytta	Lamberti and Sottile 1997, Wakasa 2008
Yagua	Payne and Payne 1990, Payne 1985
Yakan	Brainard and Behrens 2002
Yimas	Foley 1991
Yuchi	Linn 2001
Yukaghir (Kolyma)	Maslova 2003
Yuracaré	Van Gijn 2006, 2011
Zoque (Chimalapa)	Johnson 2000, Knudson 1980; **C**: Faarlund 2012a, b

Material 3: Design of the database (screenshots)

I. Rubric "Languages"; Tab "General Information"

II. Rubric "Languages"; Tab "Grammatical Profile"

III. Rubric "Complementation"; Upper part of variable catalogue

Structure and Linearization in Complex Sentence Systems
Typological database, Friedrich Schiller University Jena

Complement clause constructions

CONSTRUCTIONAL PROPERTIES

CompID	CompAwa2
CompName	CompAwaIndSpeech
LanguageName	Awa Pit

LangRefID	Curn97
CompRef	259
Status	done

CompTAM	inherent
CompVerb	converb
CompAgreement	deletion
CompCorefSubject	deletion
CompDiffSubjectP	canonical
CompDiffSubjectS	NA
CompDiffSubjectA	NA
CompModifiers	adverbial
CompNomFlag	none
CompInternalObject	canonical
CompComplexity	non-reduced
CompLinker	derivational morph
CompLinkerPos	final
CompLinkerOccur	obligatory
CompLinkerDiachr	SUB = general converb inflections ('absolute construction')

CompCommentsStructure

This is a participial/general-converb construction that can only be used with a single predicate, i.e. 'say', in order to convey indirect (rather than direct) statements. There are two participial suffixes to choose from, and each is inherently specified for a particular (relative) tense.

The construction allows same and different subjects; indexation is always deleted. Internal subject NPs are either deleted (SS) or expressed by canonical NPs (DS).

CompStructureDiachrony

ADV/COMP/(REL): the subordinators are general converbs and chiefly code underspecified adverbial relations (inference of TEMP, CAU, COND) and also form the basis for one type of relative clauses (those need to be augmented by nominalization, suggesting that we're dealing with an adverbial verb form here and not with a participle (it cannot be used by itself in relativization))

OrderPT	pre:rigid
CompOrderPT	pre:rigid
CompCentreEmbedPT	attested
CompRightExtraposPT	not attested
CompLeftExtraposPT	attested
OrderS	SV:rigid
CompOrderS	NA
CompCentreEmbedS	NA
CompExtrapositionS	NA
OrderA	initial:rigid
CompOrderA	NA
CompCentreEmbedA	NA
CompExtrapositionA	NA

CompCommentsOrderPT

This construction is fixed in preverbal position (object slot, overt statement), but the matrix subject may be extraposed beyond the main verb in order to avoid being misassigned as an argument of the subordinate clause (cf. 261)). This results in what we may here call 'left-extraposition' of the complement (for other than discourse-pragmatic reasons).

CompCommentsOrderSA

IV. Rubric "Complementation"; Lower part of variable catalogue

[…]

Material 4: *NeighborNet* analysis of argument-structural configurations in object complement clauses
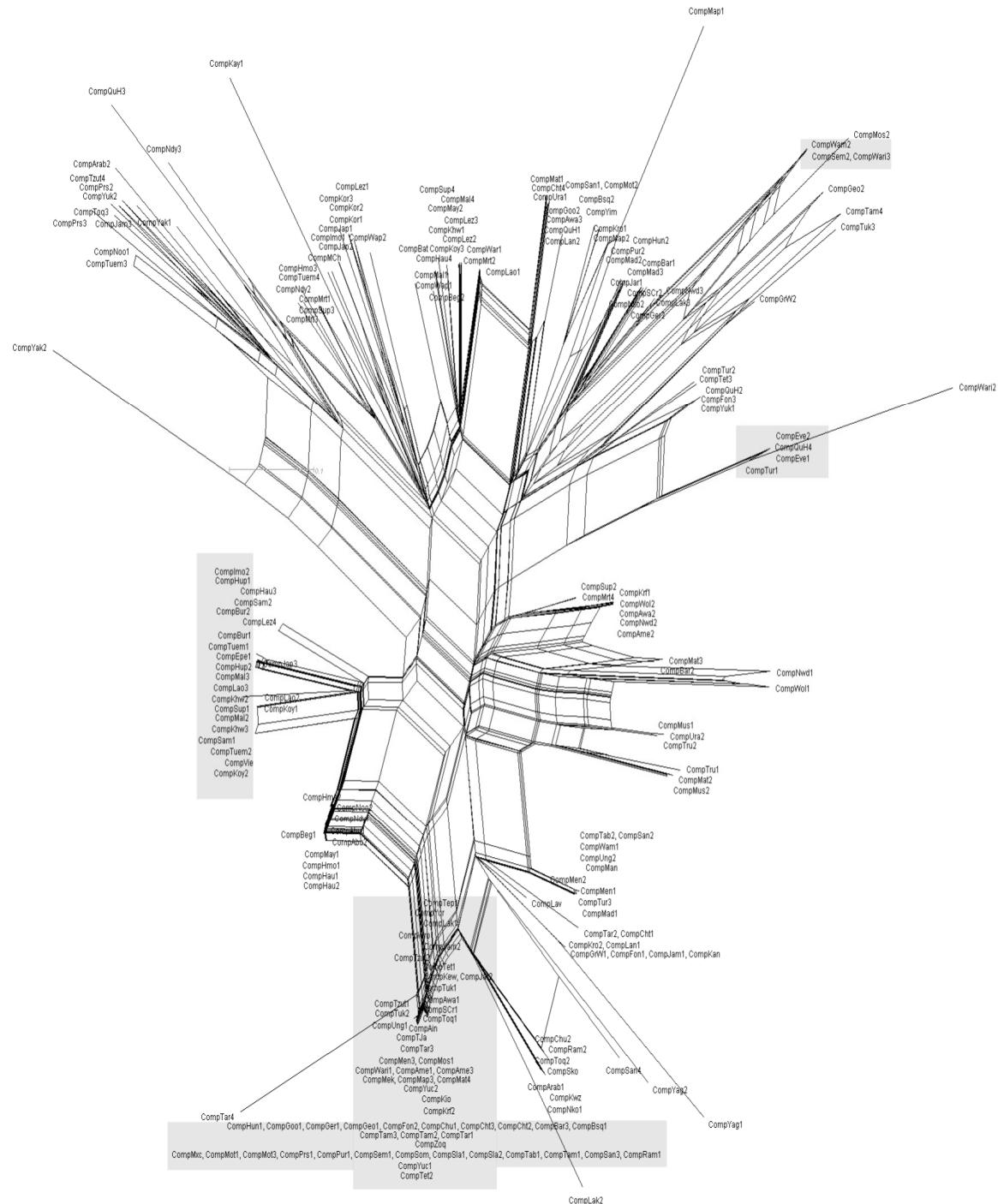
Material 5: *NeighborNet* analysis of complementation systems (I): CID mean and CID range
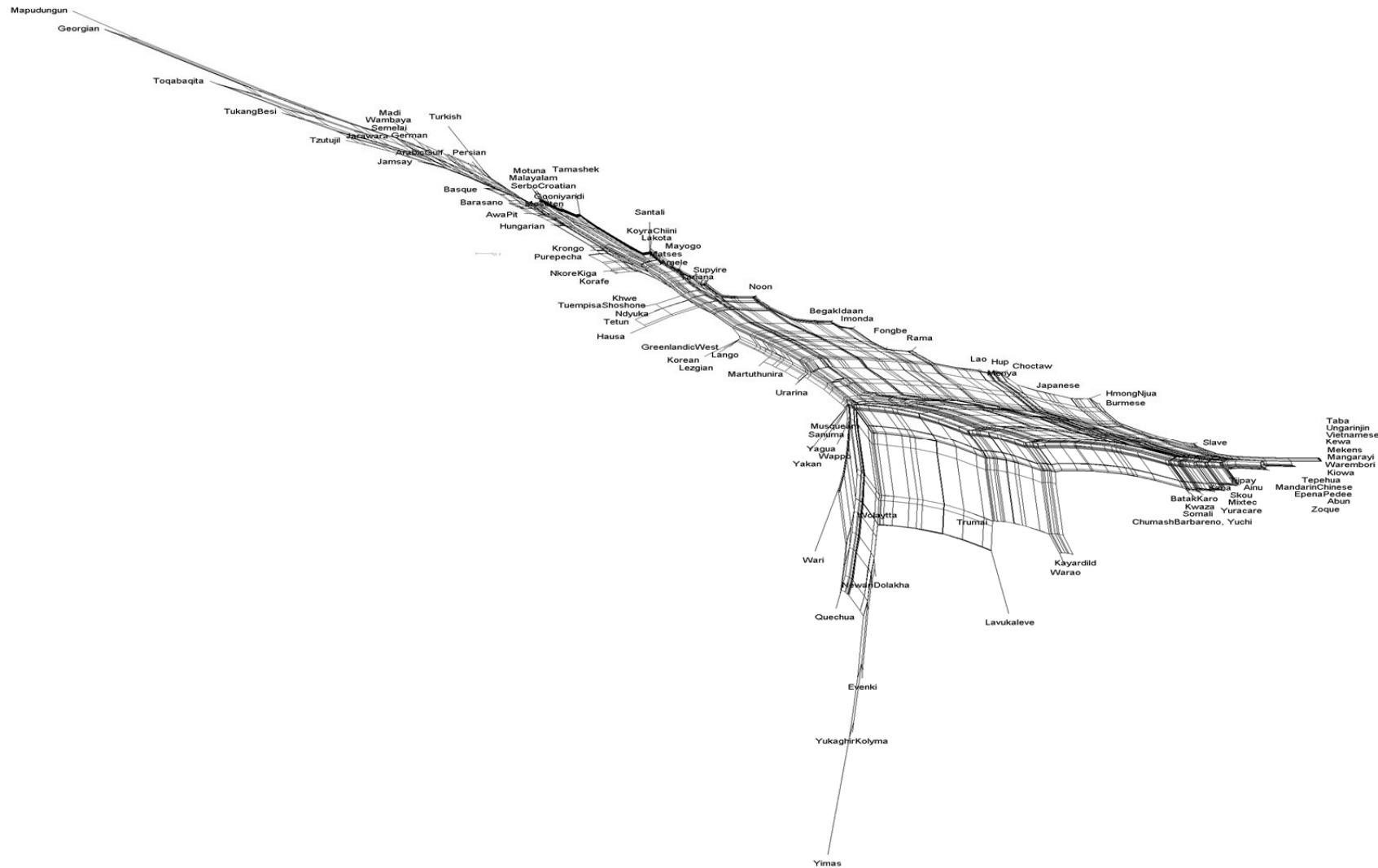
Material 6: *NeighborNet* analysis of complementation systems (II): Weighted CID mean and CID standard deviation
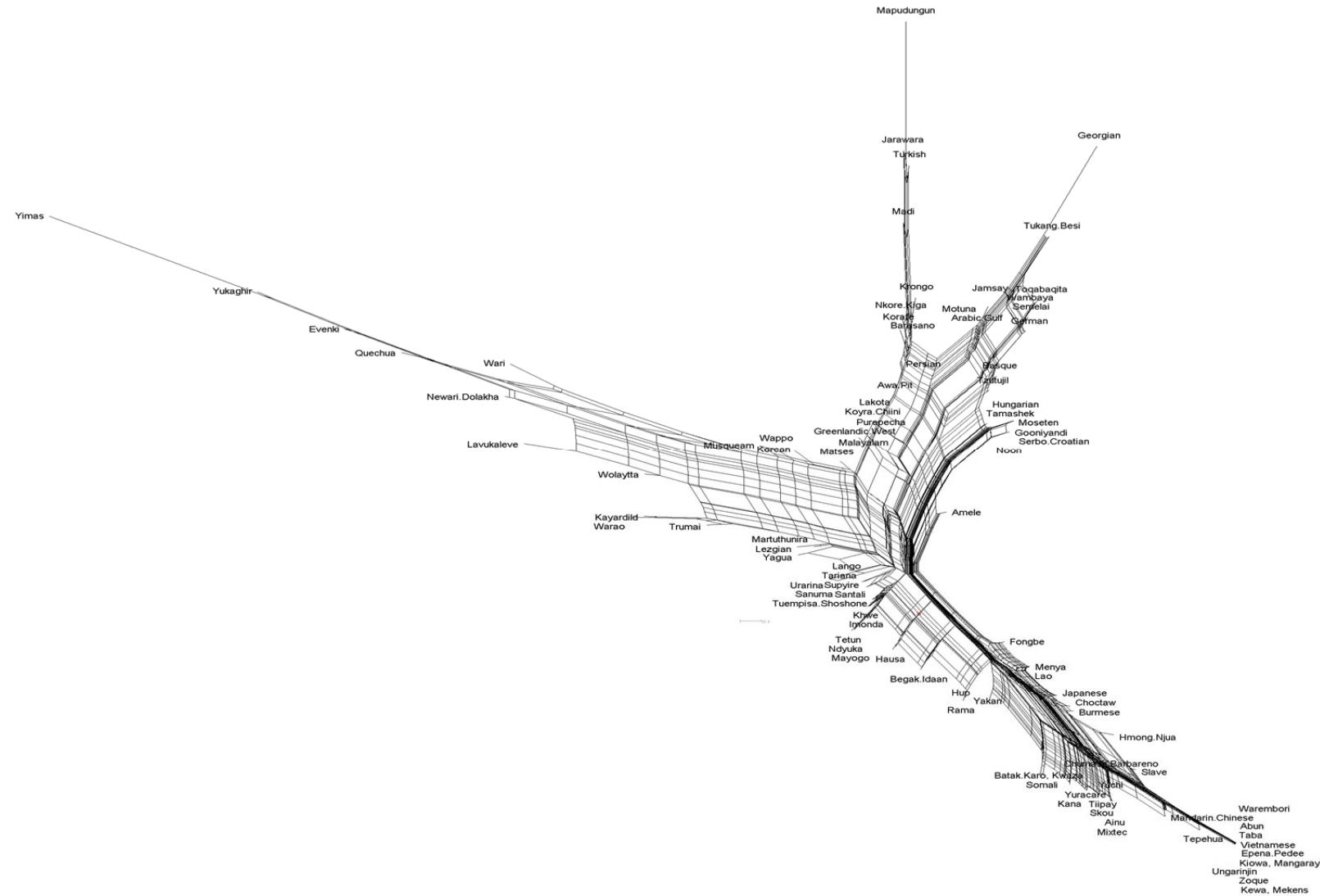
Material 7: *NeighborNet* analysis of the distributional profiles of complementation patterns (I)
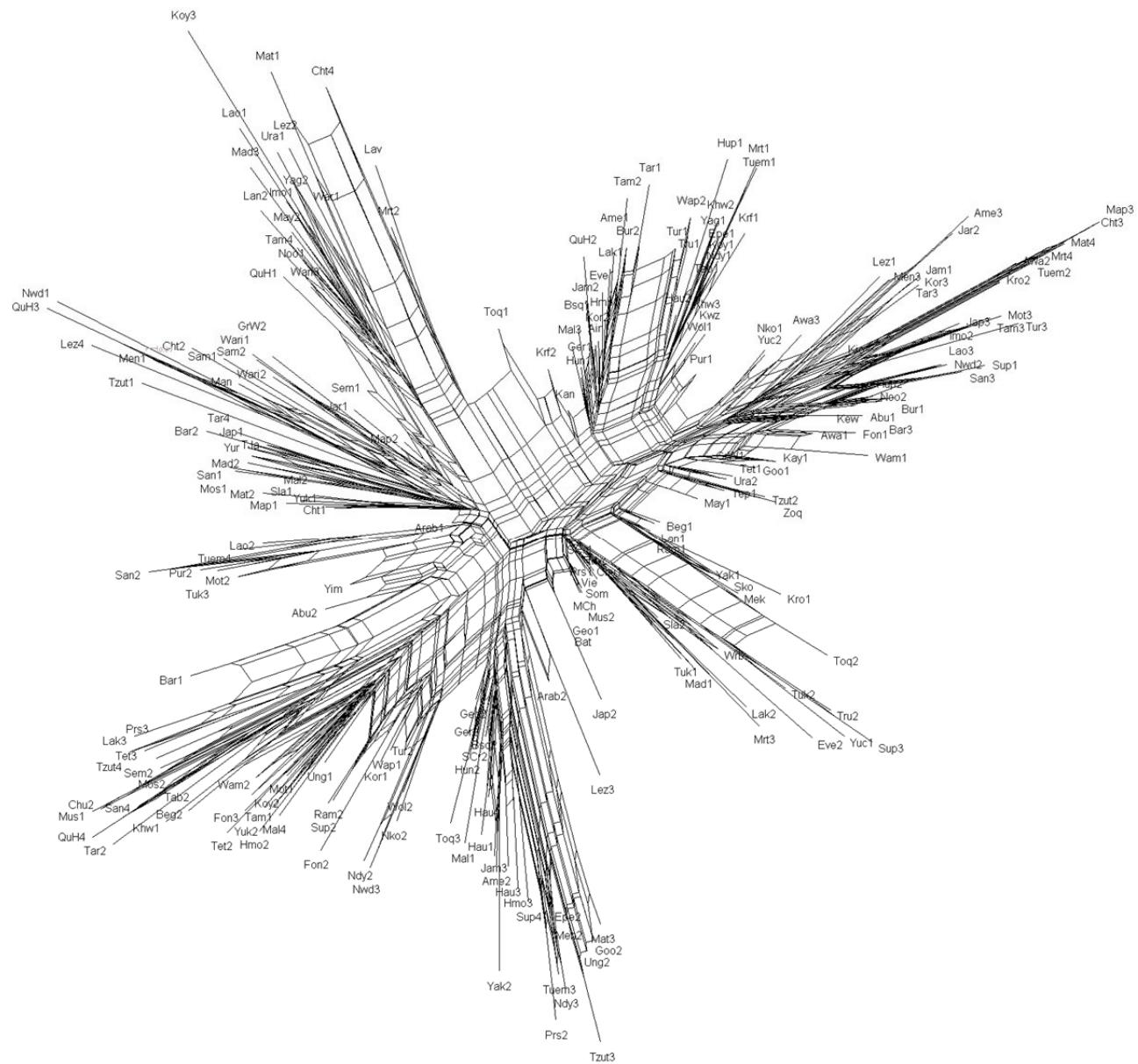
Material 8: *NeighborNet* analysis of the distributional profiles of complementation patterns (II)
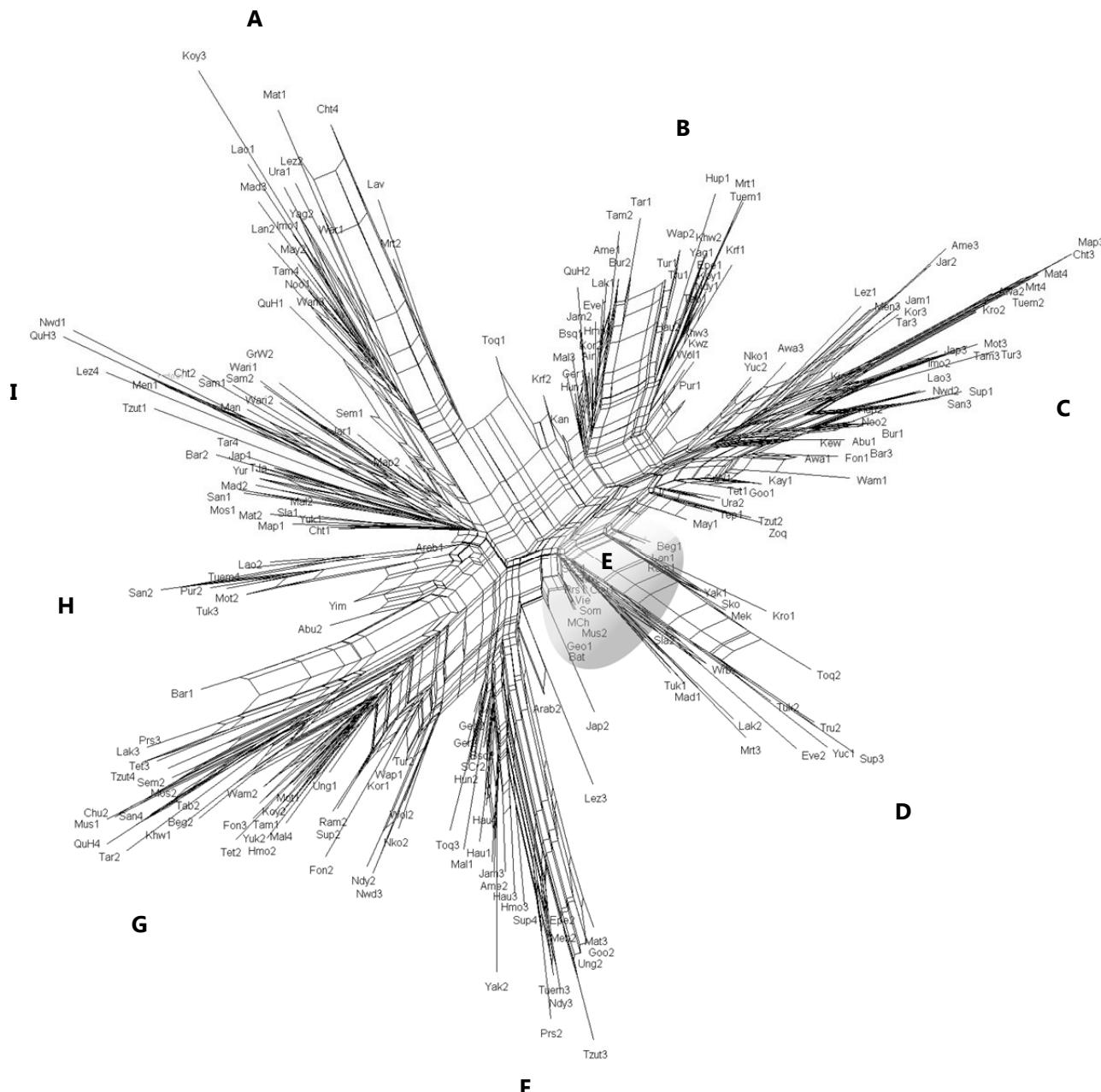
Material 9: *NeighborNet* analysis of the CTP classes according to their distributional similarities
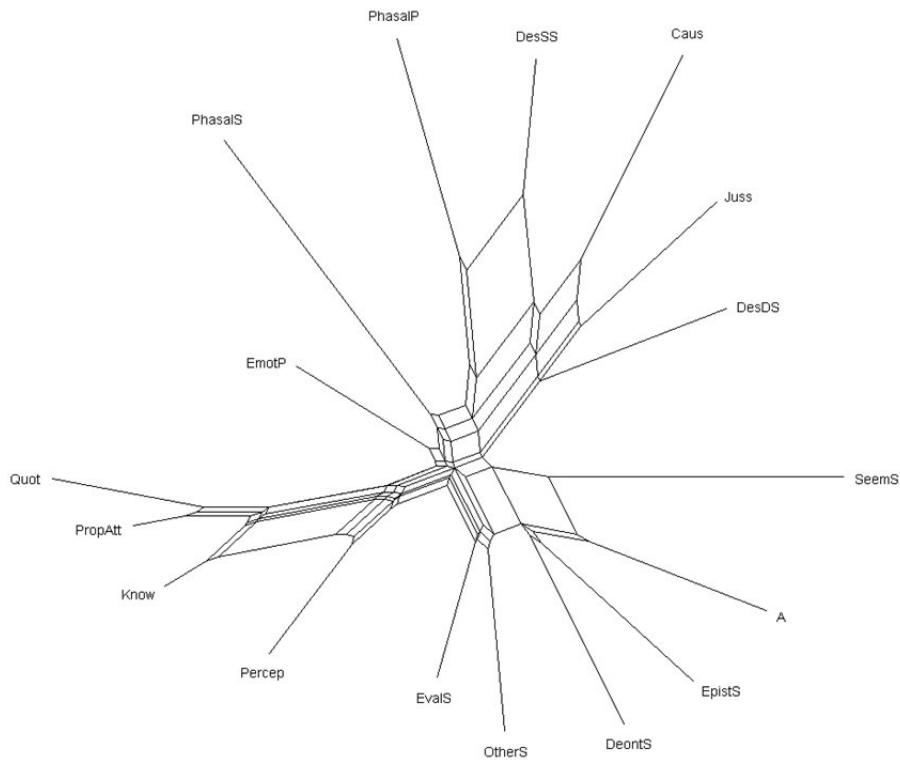
Bibliography

Abbot-Smith, Kirsten and Michael Tomasello (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review* 23: 275–290.

Abdoulaye, Mahamane L. (2008). Origin of relative marking in Hausa. Available online at *Hyper Articles en Ligne (HAL)*: <http://hal.archives-ouvertes.fr/hal-00259490/fr/> (as at July 2012).

Abdoulaye, Mahamane L. (2009). Two types of temporal *when* clauses in Hausa. *The Buckingham Journal of Language and Linguistics* 2.1: 1–19.

Abdoulaye, Mahamane L. (2011). Relative clauses in Hausa: A grammaticalization perspective. *Journal of African Languages and Linguistics* 32.1: 1–41.

Adelaar, Willem F.H. and Pieter C. Muysken (2004). *The Languages of the Andes*. Cambridge: Cambridge University Press.

Aghaei, Behrad (2006). The Syntax of *ke*-Clauses and Clausal Extraposition in Modern Persian. PhD dissertation, University of Texas at Austin.

Aikhenvald, Alexandra Y. (1999). *Tariana Texts and Cultural Context.* Munich: Lincom Europa.

Aikhenvald, Alexandra Y. (2003). *A Grammar of Tariana*. Cambridge: CUP.

Aikhenvald, Alexandra Y. (2004). *Evidentiality*. Oxford: Oxford University Press.

Aikhenvald, Alexandra Y. (2006). Complement clause types and complementation strategies in Tariana. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 178–203.

Aikhenvald, Alexandra Y. and R.M.W. Dixon (eds.) (2005). *Serial Verb Constructions: A Cross-Linguistic Typology.* Oxford: Oxford University Press.

Alba-Juez, Laura and Geoff Thompson (2014). The many facets and phases of evaluation. In: *Evaluation in Context.* Eds. Geoff Thompson and Laura Alba-Juez. Amsterdam, Philadelphia: John Benjamins. 3–23.

Aldenderfer, Mark S. and Roger K. Blashfield (1984). *Cluster Analysis.* Newbury Park, CA: Sage Publications.

Amiridze, Nino (2006). *Reflexivization in Georgian*. Utrecht: LOT Publications.

Andersen, Torben (1998). Verb-coding in Maʻdi relative clauses. *Sprachtypologie und Universalienforschung* 51.4: 295–326.

Anderson, Gregory D.S. (2006). *Auxiliary Verb Constructions.* Oxford: Oxford University Press.

Andrews, Avery D. (2007a). Major functions of the noun phrase. In: *Language Typology and Syntactic Description. Vol. I: Clause Structure.* Ed. Timothy Shopen. Cambridge: Cambridge University Press. 132–223.

Andrews, Avery D. (2007b). Relative clauses. In: *Language Typology and Syntactic Description. Vol. II: Complex Constructions.* Ed. Timothy Shopen. Cambridge: Cambridge University Press. 206–236.

Andronov, Michail S. (1996). *A Grammar of the Malayalam Language in Historical Treatment.* Wiesbaden: Harrassowitz.

Aristar, Anthony Rodrigues (1991). On diachronic sources and synchronic patterns: An investigation into the origin of linguistic universals. *Language* 67: 1–33.

Arkadiev, Peter M. (2012). Participial complementation in Lithuanian. In: *Clause Linkage in Cross-Linguistic Perspective: Data-Driven Approaches to Cross-Clausal Syntax.* Eds. Volker Gast and Holger Diessel. Berlin: Mouton de Gruyter. 285–334.

Arkadiev, Peter M. (2013). Marking of subjects and objects in Lithuanian non-finite clauses: A typological and diachronic perspective. *Linguistic Typology* 17: 397–437.

Aronson, Howard Isaac (1990). *Georgian: A Reading Grammar.* Columbus, Ohio: Slavica Publishers.

Asher, Ronald E. and T.C. Kumari (1997). *Malayalam*. London, New York: Routledge.

Asudeh, Ash and Ida Toivonen (2012). Copy raising and perception. *Natural Language and Linguistic Theory* 30.2: 321–380.

Axel, Katrin (2009). Die Entstehung des *dass*-Satzes – ein neues Szenario. In: *Koordination und Subordination im Deutschen.* Eds. Veronika Ehrich, Christian Fortmann, Ingo Reich and Marga Reis. Hamburg: Buske. 21–41.

Auer, Peter (2009). On-line syntax: Thoughts on the temporality of spoken language. *Language Sciences* 31: 1–13.

Baayen, R. Harald (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge: Cambridge University Press.

Bakker, Dik (2011). Language sampling. In: *The Oxford Handbook of Linguistic Typology.* Ed. Jae Jung Song. Oxford: Oxford University Press. 100–129.

Bagari, Dauda Muhammad (1971). Lexicalist hypothesis and Hausa. *Studies in African Linguistics* 2.3: 197–215.

Bayer, Josef (2001). Two grammars in one: Sentential complements and complementizers in Bengali and other South Asian languages. In: *The Yearbook of South Asian Languages: Tokyo Symposium on South Asian Languages - Contact, Convergence and Typology.* Eds. Peri Bhaskarorao and Karumuri V. Subbarao. New Delhi: Sage Publications. 11–36.

Beckner, Clay, Richard Blythe, Joan Bybee, Morten H. Christiansen, William Croft, Nick C. Ellis, John Holland, Jinyun Ke, Diane Larsen-Freeman and Tom Schoenemann (2009). Language is a complex adaptive system: Position paper. In: *Language as a Complex Adaptive System.* Eds. Nick C. Ellis and Diane Larsen-Freeman. Chichester: Wiley. 1–26.

Behaghel, Otto (1909). Beziehungen zwischen Umfang und Reihenfolge von Satzgliedern. *Indogermanische Forschungen* 25: 110–142.

Belić, Bojan (2005). Complement Verb Variation in Present-Day Serbian. PhD disseration, Ohio State University.

Belletti, Adriana and Luigi Rizzi (1988). Psych-verbs and θ-theory. *Natural Language and Linguistic Theory* 6: 291–352.

Benkő, Loránd (ed.) (1993). *Etymologisches Wörterbuch des Ungarischen.* Budapest: Akad. Kiadó

Berchem, Jörg (1991). *Referenzgrammatik des Somali.* Köln: OMIMEE.

Bergh, Gunnar (1997). Vacuous extraposition from object in English. *Studia Neophilologica* 69: 37–41.

Berman, Arlene (1974). Adjectives and Adjective Complement Constructions in English. PhD dissertation, Harvard University.

Berry, Christine (1995). Complex Sentences in the Abun Language. MA thesis, La Trobe University.

Berry, Keith and Christine Berry (1999). *A Description of Abun: a West Papuan Language of Irian Jaya.* Canberra: Pacific Linguistics.

Bickel, Balthasar (2003). Referential density in discourse and in syntactic typology. *Language* 79.4: 708–736.

Bickel, Balthasar (2004). The syntax of experiencers in the Himalayas. In: *Non-Nominative Subjects. Vol. I.* Eds. Peri Bhaskararao and Karumuri Venkata Subbarao. Amsterdam, Philadelphia: John Benjamins. 77–111.

Bickel, Balthasar (2007). Typology in the 21st century: Major current developments. *Linguistic Typology* 11: 239–251.

Bickel, Balthasar (2008a). A refined sampling procedure for genealogical control. *Sprachtypologie und Universalienforschung* 61: 221–233.

Bickel, Balthasar (2008b). A general method for the statistical evaluation of typological distributions. Ms., University of Leipzig. Available online at <http://www.spw.uzh.ch/bickel/publications/ manuscripts_en.html> (as at September 2013).

Bickel, Balthasar (2010). Capturing particulars and universals in clause linkage: A multivariate analysis. In: *Clause Linking and Clause Hierarchy: Syntax and Pragmatics.* Ed. Isabelle Bril. Amsterdam, Philadelphia: John Benjamins. 51–102.

Bickel, Balthasar (2011). Grammatical relations typology. In: *The Oxford Handbook of Linguistic Typology.* Ed. Jae Jung Song. Oxford: Oxford University Press. 399–444.

Bickel, Balthasar and Johanna Nichols (2002). Autotypologizing databases and their Uuse in fieldwork. In: *Proceedings of the International LREC Workshop on Resources and Tools in Field Linguistics, Las*

Palmas, 26–27 May 2002. Eds. Peter Austin, Helen Dry and Peter Wittenburg. Nijmegen: ISLE and DOBES.

Bickel, Balthasar and Johanna Nichols (2006). Oceania, the Pacific Rim, and the theory of linguistic areas. *Proceedings of the 32nd Annual Meeting of the Berkeley Linguistics Society* 32S: 3–15.

Bisang, Walter (1992). *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer.* Tübingen: Narr.

Blackings, Mairi and Nigel Fabb (2003). *A Grammar of Ma'di.* Berlin, New York: Mouton de Gruyter.

Bock, Hellmut (1931). Studien zum präpositionalen Infinitiv und Akkusativ mit dem *to*-Infinitiv. *Anglia* 55: 114–249.

Bolinger, Dwight (1977). *Meaning and Form.* London: Longman.

Bolinger, Dwight (1980). *Wanna* and the gradience of auxiliaries. In: *Wege zur Universalienforschung: Sprachwissenschaftliche Beiträge zum 60. Geburtstag von Hansjakob Seiler.* Eds. Gunter Brettschneider and Christian Lehmann. Tübingen: Narr. 292–299.

Borgman, Donald M. (1990). Sanuma. In: *Handbook of Amazonian Languages. Vol. 2.* Eds. Desmond C. Derbyshire and Geoffrey K. Pullum. Berlin, New York: Mouton de Gruyter. 15–248.

Bornkessel-Schlesewsky, Ina, Kamal Kumar Choudhary, Alena Witzlack-Makarevich and Balthasar Bickel (2008). Bridging the gap between processing preferences and typological distributions. In *Scales.* Eds. Andrej Malchukov and Marc Richards. Linguistische ArbeitsBerichte 86. University of Leipzig. 397–436.

Bossong, Georg (1979). Typologie der Hypotaxe. *Folia Linguistica* 13: 33–54.

Bossong, Georg (1998). Le marquage de l'expérient dans les langues d'Europe. In: *Actance et Valence dans les Langues de l'Europe.* Ed. Jacques Feuillet. Berlin: Mouton de Gruyter. 259–294.

Bowden, John (1997). Taba (Makian Dalam). Description of an Austronesian Language from Eastern Indonesia. PhD dissertation, University of Melbourne.

Bowden, John (2001). *Taba.* Canberra: Pacific Linguistics.

Boye, Kasper (2010). Reference and clausal perception-verb complements. *Linguistics* 48.2: 391–430.

Brainard, Sherry and Dietlinde Behrens (2002). *A Grammar of Yakan.* Manila: Linguistic Society of the Philippines.

Bresnan, Joan, Shipra Dingare and Christopher D. Manning (2001). Soft constraints mirror hard constraints: Voice and person in English and Lummi. In: *Proceedings of the LFG01 Conference.* Available online at <http://cslipublications.stanford.edu/LFG/6/lfg01-toc.html> (as at January 2007).

Broadwell, George Aaron (2006). *A Choctaw Reference Grammar.* Lincoln, London: University of Nebraska Press.

Brockelmann, Carl (1908). *Vergleichende Grammatik der semitischen Sprachen: Elemente der Laut- und Formenlehre.* Berlin: Verlag von Reuther und Reichard.

Browne, Wayles (1987). Classification of subordinate clauses in a grammar of Serbo-Croatian for foreign users. In: *Contrastive Analysis of English and Serbo-Croatian. Vol. III: Verbal Tenses.* Ed. Rudlof Filipović. Zagreb: Instititue of Linguistics, University of Zagreb. 165–191.

Browne, Wayles (1993). Serbo-Croat. In: *The Slavonic Languages.* Eds. Bernard Comrie and Greville Corbett. London, New York: Routledge. 306–387.

Brownes, Wayles and Theresa Alt (2004). *A Handbook of Bosnian, Serbian and Croatian.* Duke University and the University of North Carolina at Chapel Hill: SEELRC.

Bryant, David and Vincent Moulton (2004). Neighbor-Net: An agglomerative method for the construction of phylogenetic networks. *Molecular Biology and Evolution* 21.2: 255–265.

Buechel, Eugene (1939). *A Grammar of Lakota: The Language of the Teton Sioux Indians.* St. Louis, Chicago, New York, Cincinnati: John S. Swift Co., Inc.

Bugaeva, Anna (2008a). Reported discourse and logophoricity in Southern Hokkaido dialects of Ainu. *Gengo Kenkyu* 133: 31–75.

Bugaeva, Anna (2008b). Speech report constructions in Ainu. In: *Subordination and Coordination Strategies in North Asian Languages.* Ed. Edward J. Vajda. Amsterdam, Philadelphia: John Benjamins. 17–29.

Bugaeva, Anna (2012). Southern Hokkaido Ainu. In: *The Languages of Japan and Korea.* Ed. Nicolas Tranter. London, New York: Routledge. 461–509.

Bybee, Joan L. (1985). *Morphology: A Study of the Relation between Meaning and Form.* Amsterdam, Philadelphia: John Benjamins.

Bybee, Joan (2006). From usage to grammar: The mind's response to repetition. *Language* 82.4: 711–733.

Bybee, Joan (2010). *Language, Usage and Cognition.* Cambridge: Cambridge University Press.

Bybee, Joan, Revere Perkins and William Pagliuca (1994). *The Evolution of Grammar: Tense, Aspect, and Modality in the Languages of the World.* Chicago: Chicago University Press.

Bybee, Joan L. and Sandra Thompson (1997). Three frequency effects in syntax. *Berkeley Linguistics Society* 23: 378–388.

Campbell, Lyle, Terrence Kaufman and Thomas C. Smith-Stark (1986). Meso-America as a linguistic area. *Language* 62.3: 530–570.

Carlson, Robert (1994). *A Grammar of Supyire.* Berlin, New York: Mouton de Gruyter.

Chamereau, Claudine (2000). *Grammaire du Purépecha Parlé sur des Îles du Lac de Patzcuaro.* München: Lincom Europa.

Chappell, Hillary (2008). Variation in the grammaticalization of complementizers from *verbal dicendi* in Sinitic languages. *Linguistic Typology* 12.1: 45–98.

Chung, Chan (1995). A Lexical Approach to Word Order Variation in Korean. PhD dissertation, The Ohio State University.

Clark, Herbert H. and Richard J. Gerrig (1990). Quotations as demonstrations. *Language* 66.4: 764–805.

Clark, Marybeth (1996). Conjunction as copula in Vietnamese. *Mon-Khmer Studies* 26: 319–331.

Cole, Peter, Wayne Harbert, Gabriella Hermon and S.N. Sridhar (1980). The acquisition of subjecthood. *Language* 56: 719–743.

Comrie, Bernard (1976). The syntax of action nominalizations: A cross-language study. *Lingua* 40: 177–201.

Comrie, Bernard (1989). *Language Universals and Linguistic Typology.* 2nd ed. Chicago: The University of Chicago Press.

Comrie, Bernard (1993). Argument structure. In: *Syntax: An International Handbook. Vol. I.* Eds. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann. Berlin, New York: Walter de Gruyter. 905–914.

Comrie, Bernard (1996). The unity of noun-modifying clauses in Asian languages. In: *Pan-Asiatic Linguistics: Proceedings of the Fourth International Symposium on Languages and Linguistics*, January 8–10, 1996. Salaya, Thailand: Institute of Language and Culture for Rural Development, Mahidol University at Salaya. 1077–1088.

Comrie, Bernard (2011). Alignment of case marking of full noun phrases. In: *The World Atlas of Language Structures Online.* Eds. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, chapter 98. Available online at <http://wals.info/chapter/98> (as at May 2012).

Comrie, Bernard and Zarina Estrada-Fernández (eds.) (2012). *Relative Clauses in the Languages of the Americas.* Amsterdam, Philadelphia: John Benjamins.

Comrie, Bernard and Stephen Matthews (1990). Prolegomena to a typology of Tough Movement. In: *Studies in Typology and Diachrony: Papers Presented to Joseph H. Greenberg on his 75th Birthday.* Eds. William Croft, Keith Denning and Suzanne Kemmer. Amsterdam, Philadelphia: John Benjamins. 43–58.

Comrie, Bernard and Sandra A. Thompson (2007) [1985]. Lexical nominalization. In: *Language Typology and Syntactic Description. Vol. III: Grammatical Categories and the Lexicon.* Ed. Timothy Shopen. Cambridge: Cambridge University Press. 334–381.

Cooper, David L. (1999). *Linguistic Attractors: The Cognitive Dynamics of Language Acquisitiona and Change.* Amsterdam, Philadelphia: John Benjamins.

Coulmas, Florian (1986). Direct and indirect speech in Japanese. In: *Direct and Indirect Speech.* Ed. Florian Coulmas. Berlin, New York: Mouton de Gruyter. 161–178.

Couper-Kuhlen, Elizabeth and Sandra A. Thompson (2006). "You know it's funny." Eine Neubetrachtung der 'Extraposition' im Englischen. In: *Konstruktionen in der Interaktion.* Eds. Susanne Günthner and Wolfgang Imo. Berlin, New York: de Gruyter. 23–58.

Creissels, Denis, Gerrit J. Dimmendaal, Zgymunt Frajzyngier and Christa König (2008). Africa as a morphosyntactic area. In: *A Linguistic Geography of Africa.* Eds. Bernd Heine and Derek Nurse. Cambridge: Cambridge University Press. 6–150.

Cristofaro, Sonia (2003). *Subordination.* Oxford: Oxford University Press.

Cristofaro, Sonia (2008). A constructionist approach to complementation: Evidence from Ancient Greek. *Linguistics* 46.3: 571–606.

Croft, William (1993). Case marking and the semantics of mental verbs. In: *Semantics and the Lexicon.* Ed. James Pustejovsky. Dordrecht etc.: Kluwer. 55–72.

Croft, William (2000). *Explaining Language Change: An Evolutionary Approach*. Harlow: Longman.

Croft, William (2001). *Radical Construction Grammar: Syntactic Theory in Typological Perspective.* Oxford: Oxford University Press.

Croft, William (2003). *Typology and Universals.* 2nd ed. Cambridge: Cambridge University Press.

Croft, William (2007). The origins of grammar in the verbalization of experience. *Cognitive Linguistics* 18.3: 339–382.

Croft, William (2012). *Verbs: Aspect and Causal Structure.* Oxford: Oxford University Press.

Croft, William and Keith T. Poole (2008). Inferring Universals from Grammatical Variation: Multidimensional Scaling for Typological Analysis. *Theoretical Linguistics* 34.1: 1–37.

Csató, Éva Á. (2010). Two types of complement clauses in Turkish. In: *Turcology in Mainz.* Eds. Hendrik Boeschoten and Julian Rentzsch. Wiesbaden: Harrassowitz. 107–122.

Culy, Christopher D. (1990). The Syntax and Semantics of Internally Headed Relative Clauses. PhD dissertation, Stanford University.

Culy, Christopher D. (1994). Aspects of logophoric marking. *Linguistics* 32.6: 1055–1094.

Curnow, Timothy J. (1997). A Grammar of Awa Pit (Cuaiquer). PhD dissertation, Australian National University.

Curnow, Timothy J. (2002). Conjunct/disjunct marking in Awa Pit. *Linguistics* 40.3: 611–627.

Cysouw, Michael (2005). Quantitative methods in typology. In: *Quantitative Linguistics: An International Handbook.* Eds. Reinhard Köhler, Gabriel Altmann and Rajmund G. Piotrowski. Berlin, New York: de Gruyter. 554–578.

Cysouw, Michael (2007). New approaches to cluster analysis of typological indices. In: *Exact Methods in the Study of Language and Text.* Eds. Peter Grzybek and Reinhard Köhler. Berlin, New York: Mouton de Gruyter. 61–76.

Dąbrowska, Ewa (2004). *Language, Mind and Brain: Some Psychological and Neurological Constraints on Theories of Grammar.* Washington, D.C.: Georgetown University Press.

Dąbrowska, Ewa and Elena Lieven (2005). Towards a lexically specific grammar of children's question constructions. *Cognitive Linguistics* 16.3: 437–474.

Dahl, Östen and Kari Fraurud (1996). Animacy in grammar and discourse. In: *Reference and Referent Accessibility*. Eds. Thorstein Fretheim and Jeanette K. Gundel. Amsterdam, Philadelphia: John Benjamins. 47–64.

Davies, William D. (1981). Choctaw Clause Structure. PhD dissertation, University of San Diego.

Dayley, Jon P. (1985). *Tzutujil Grammar.* Berkeley: University of California Press.

Dayley, Jon P. (1989). *Tümpisa (Panamint) Shoshone Grammar.* Berkeley: University of California Press.

de Groot, Casper (1995). The Hungarian converb or verbal adverbial in *-va/-ve*. In: *Converbs in Cross-Linguistic Perspective.* Eds. Martin Haspelmath and Ekkehard König. Berlin, New York: Mouton de Gruyter. 283–311.

de Haan, Ferdinand (2006). Typological approaches to modality. In: *The Expression of Modality.* Eds. William Frawley, Erin Eschenroeder, Sarah Mills and Thao Nguyen. Berlin, New York: Mouton de Gruyter. 27–69.

de Roeck, Marijke (1994). A functional typology of speech reports. In: *Function and Expression in Functional Grammar.* Eds. Elisabeth Engberg-Pedersen, Lisbeth Falster Jakobsen and Lone Schack Rasmussen. Berlin, New York: Mouton de Gruyter. 331–351.

de Smedt, Koenraad J.M.J. (1994). Parallelism in incremental sentence generation. In: *Parallelism in Natural Language Processing.* Eds. Geert Adriaens and Udo Hahn. Norwood, NJ: Ablex. 421–448.

de Smet, Hendrik (2012). The course of actualization. *Language* 88.3: 601–633.

Dehghani, Yavar (n.d.). Causative constructions in Persian and Azari. Ms. Available online at < http://www.iranianlinguistics.org/papers/05Dehghani.pdf> (as at May 2012).

DeLancey, Scott (1981). An interpretation of split ergativity and related patterns. *Language* 57.3: 626–657.

DeLancey, Scott (1997). Mirativity: The grammatical marking of unexpected information. *Linguistic Typology* 1: 33–52.

Dench, Alan C. (1995). *Martuthunira.* Canberra: Pacific Linguistics.

Dench, Alan and Nicholas Evans (1988). Multiple case-marking in Australian languages. *Australian Journal of Linguistics* 8: 1–47.

Deutscher, Guy (2000). *Syntactic Change in Akkadian: The Evolution of Sentential Complementation.* Oxford: Oxford University Press.

Deutscher, Guy (2006). Complement clause types and complementation strategies in Akkadian. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 159–177.

Deutscher, Guy (2011). The grammaticalization of quotatives. In: *The Oxford Handbook of Grammaticalization.* Eds. Heiko Narrog and Bernd Heine. Oxford: Oxford University Press. 646–655.

Diessel, Holger (2001). The ordering distribution of main and adverbial clauses: A typological study. *Language* 77.2: 433–455.

Diessel, Holger (2008). Iconicity of sequence. A corpus-based analysis of the positioning of temporal adverbial clauses in English. *Cognitive Linguistics* 19: 457–482.

Diessel, Holger (2011). Review article of *Language, Usage and Cognition* by Joan Bybee. *Language* 87: 830–844.

Diessel, Holger and Katja Hetterle (2011). Causal clauses: A cross-linguistic investigation of their structure, meaning, and use. In: *Linguistic Universals and Language Variation.* Ed. Peter Siemund. Berlin, New York: Mouton de Gruyter. 21–52.

Dik, Simon C. (1997). *The Theory of Functional Grammar. Part I: The Structure of the Clause.* Berlin, New York: Mouton de Gruyter.

Dik, Simon C. and Kees Hengeveld (1991). The hierarchical structure of the clause and the typology of perception-verb complements. *Linguistics* 29: 231–259.

Dimitriadis, Alexis and Simon Musgrave (2009). Designing linguistic databases: A primer for linguists. In: *The Use of Databases in Cross-Linguistic Studies.* Eds. Martin Everaert, Simon Musgrave and Alexis Dimitriadis. Berlin, New York: Mouton de Gruyter. 13–75.

Disterheft, Dorothy (1980). *The Syntactic Development of the Infinitive in Indo-European.* Columbus, Ohio: Slavica Press.

Dixon, R.M.W. (1995). Complement clauses and complementation strategies. In: *Grammar and Meaning. Essays in Honour of Sir John Lyons.* Ed. Frank R. Palmer. Cambridge: Cambridge University Press. 175–220.

Dixon, R.M.W. (2002). *Australian Languages: Their Nature and Development.* Cambridge: Cambridge University Press.

Dixon, R.M.W. (2004). *The Jarawara Language of Southern Amazonia.* Oxford: Oxford University Press.

Dixon, R.M.W. (2005). *A Semantic Approach to English Grammar.* Oxford: Oxford University Press.

Dixon, R.M.W. (2006a). Complement clauses and complementation strategies in typological perspective. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 1–48.

Dixon, R.M.W. (2006b). Complement clause type and complementation strategy in Jarawara. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 93–114.

Dixon, R.M.W. (2006c). Complementation strategies in Dyirbal. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 263–279.

Dixon, R.M.W. and Alexandra Y. Aikhenvald (eds.) (2006). *Complementation: A Cross-Linguistic Typology.* Oxford: Oxford University Press.

Dol, Philomena (1999). A Grammar of Maybrat: A Language of the Bird's Head, Irian Jaya. PhD dissertation, Leiden University.

Donner, Morton (1986). The Gerund in Middle English. *English Studies* 67: 394–400.

Donohue, Mark (1999a). *Warembori.* München: Lincom.

Donohue, Mark (1999b). *A Grammar of Tukang Besi.* Berlin, New York: Mouton de Gruyter.

Donohue, Mark (2004). A Grammar of the Skou Language of New Guinea. Draft version. Available online at <http://papuan.linguistics.anu.edu.au/Skou.php> (as at February 2008).

Donohue, Mark (2008). Semantic alignment systems: what's what, and what's not. In: *The Typology of Semantic Alignment.* Eds. Mark Donohue and Søren Wichmann. Oxford: Oxford University Press. 24–75.

Dörnyei, Zoltán (2009). *The Psychology of Second Language Acquisition.* Oxford: Oxford University Press.

Dryer, Matthew S. (1980). The positional tendencies of sentential noun phrases in universal grammar. *Canadian Journal of Linguistics* 25.2: 123–195.

Dryer, Matthew S. (1986). Primary objects, secondary objects, and antidative. *Language* 62: 808–845.

Dryer, Matthew S. (1989a). Large linguistic areas and language sampling. *Studies in Language* 13.2: 257–292.

Dryer, Matthew S. (1989b). Discourse-governed word order and word order typology. *Belgian Journal of Linguistics* 4: 69–90.

Dryer, Matthew S. (1997). On the six-way word order typology. *Studies in Language* 21: 69–103.

Dryer, Matthew S. (2003). Word order in Sino-Tibetan languages from a typological and geographical perspective. In: *The Sino-Tibetan Languages.* Eds. Graham Thurgood and Randy J. LaPolla. London: Routledge. 43–56.

Dryer, Matthew S. (2007). Clause types. In: *Language Typology and Syntactic Description. Vol. I: Clause Structure.* Ed. Timothy Shopen. Cambridge: Cambridge University Press. 224–275.

Dryer, Matthew S. (2009). The branching direction theory of word order correlations revisited. In: *Universals of Language Today.* Eds. Sergio Scalise, Elisabetta Magni and Antionietta Bisetto. Dordrecht: Springer. 185–208.

Dryer, Matthew S. (2011a). Expression of pronominal subjects. In: *The World Atlas of Language Structures Online.* Eds. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, chapter 101. Available online at <http://wals.info/chapter/101> (as at May 2012).

Dryer, Matthew S. (2011b). Order of relative clause and noun. In: *The World Atlas of Language Structures Online.* Eds. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, chapter 90A. Available online at <http://wals.info/chapter/90A> (as at May 2012).

Dryer, Matthew S. (2011c). Order of subject, object and verb. In: *The World Atlas of Language Structures Online.* Eds. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, chapter 81. Available online at <http://wals.info/chapter/81> (as at May 2012).

Dryer, Matthew S. (2011d). Order of subject and verb. In: *The World Atlas of Language Structures Online.* Eds. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, chapter 82. Available online at <http://wals.info/chapter/82> (as at May 2012).

Dryer, Matthew S. (2011e). Order of object and verb. In: *The World Atlas of Language Structures Online.* Eds. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, chapter 83A. Available online at <http://wals.info/chapter/83A> (as at May 2012).

Dryer, Matthew S. and Martin Haspelmath (eds.) (2011). *The World Atlas of Language Structures Online.* Munich: Max Planck Digital Library. Available online at <http://wals.info/> (as at May 2012).

DuBois, John A. (1987). The discourse basis of ergativity. *Language* 64: 805–855.

Ebert, Karen H. (1991). Vom Verbum dicendi zur Konjunktion: ein Kapitel universaler Grammatikentwicklung. In: *Von Europa bis Ozeanian: Von der Antonymie zum Relativsatz. Gedenkschrift für Meinrad Scheller.* Eds. Walter Bisang and Peter Rinderknecht. Zürich: Universität Zürich. 77–95.

Eisenberg, Peter (2006). *Grundriss der deutschen Grammatik. Band 2: Der Satz.* 3rd ed. Stuttgart, Weimar: Metzler.

Emonds, Joseph (1972). A reformulation of certain syntactic transformations. In: *Goals of Linguistic Theory.* Ed. Stanley Peters. Englewood Cliffs, NJ: Prentice-Hall. 21–62.

Enfield, Nick J. (2007). *A Grammar of Lao.* Berlin, New York: Mouton de Gruyter.

Englebretson, Robert (2003). *Searching for Structure: The Problem of Complementation in Colloquial Indonesian Conversation.* Amsterdam, Philadelphia: John Benjamins.

Epps, Patience (2008). *A Grammar of Hup.* Berlin, New York: Mouton de Gruyter.

Erdal, Marcel (2004). *A Grammar of Old Turkic.* Leiden, Boston: Brill.

Erdmann, Oskar (1874). *Untersuchungen über die Syntax der Sprache Otfrids.* Halle. (Reprint: Hildesheim, New York: Georg Olms Verlag, 1973)

Erdmann, Peter (1987). It-*Sätze im Englischen.* Heidelberg: Winter.

Ersen-Rasch, Margarete I. (2004). *Türkische Grammatik.* Ismaning: Hueber.

Evans, Nicholas (1995). *A Grammar of Kayardild.* Berlin, New York: Mouton de Gruyter.

Evans, Nicholas (2004). Experiencer objects in Iwaidjan languages (Australia). In: *Non-Nominative Subjects. Vol. I.* Eds. Peri Bhaskararao and Karumuri Venkata Subbarao. Amsterdam, Philadelphia: John Benjamins. 169–192.

Everett, Caleb (2008). The origins of preferred argument structure. In: *Proceedings of the 7ᵗʰ International Conference on the Evolution of Language.* Eds. Andrew D.M. Smith, Kenny Smith and Ramon Ferrer i Cancho. Hackensack, NJ: World Scientific. 107–114.

Everett, Daniel L. and Barbara Kern (1997). *Wari' – The Pacaas Novos Language of Western Brazil.* London: Routledge.

Faarlund, Jan Terje (2012a). Degrees of clause cohesion: Complementation and subordination in Chiapas Zoque. *Linguistic Discovery* 10.1: 62–74.

Faarlund, Jan Terje (2012b). *A Grammar of Chiapas Zoque.* Oxford: Oxford University Press.

Fabricius-Hansen, Cathrine (2005). Das Verb. In: *Duden: Die Grammatik.* Eds. Matthias Wermke, Kathrin Kunkel-Razum and Wener Scholze-Stubenrecht. Mannheim etc.: Dudenverlag. 395–572.

Falk, Yehuda N. (2006). *Subjects and Universal Grammar: An Explanatory Theory.* Cambridge: Cambridge University Press.

Fanego, Teresa (2004). On reanalysis and actualization in syntactic change: The rise and development of English verbal gerunds. *Diachronica* 21.1: 5–55.

Fanego, Teresa (2007). Drift and the development of sentential complements in British and American English from 1700 to the present day. In: *Of Varying Language and Opposing Creed: New Insights into Late Modern English.* Eds. Dolores González-Álvarez, Jorges L. Bueno-Alonso and Esperanza Rama-Martínez. Bern: Peter Lang. 161–235.

Farr, Cynthia J.M. (1999). *The Interface between Syntax and Discourse in Korafe.* Canberra: Pacific Linguistics.

Field, Andy, Jeremy Miles and Zoë Field (2012). *Discovering Statistics Using R.* Los Angeles etc.: Sage Publications.

Fleck, David W. (2003). A Grammar of Matsés. PhD dissertation, Rice University.

Fleck, David W. (2006). Complement clause type and complementation strategies in Matses. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 224–244.

Fischer, Olga (2007). *Morphosyntactic Change: Functional and Formal Perspectives.* Oxford: Oxford University Press.

Fischer, Olga and Wim van der Wurff (2006). Syntax. In: *A History of the English Language.* Eds. Richard Hogg and David Denison. Cambridge: Cambridge University Press. 109–198.

Fleck, David W. (2003). A Grammar of Matsés. PhD dissertation, Rice University.

Fleck, David W. (2006). Complement clause type and complementation strategies in Matses. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 224–244.

Foley, William A. (1986). *The Papuan Languages of New Guinea.* Cambridge: Cambridge University Press.

Foley, William A. (1991). *The Yimas Language of New Guinea.* Stanford: Stanford University Press.

Foley, William A. and Robert D. VanValin Jr. (1984). *Functional Syntax and Universal Grammar.* Cambridge: Cambridge University Press.

Ford, Cecilia E. and Sandra A. Thompson (1996). Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the projection of turn completion. In: *Interaction and Grammar.* Eds. Elinor Ochs, Emanuel A. Schegloff and Sandra A. Thompson. Cambridge: Cambridge University Press. 135-184.

Ford, Cecilia E., Barbara A. Fox and Sandra A. Thompson (2002). Constituency and the grammar of turn increments. In: *The Language of Turn and Sequence.* Eds. Cecilia E. Ford, Barbara A. Fox and Sandra A. Thompson. New York: Oxford University Press. 14–38.

Fortescue, Micheal (1984). *West Greenlandic.* London: Croom Helm.

Foster, Mary Lecron (1969). *The Tarascan Language.* Berkeley and Los Angeles: University of California Press.

Frajzyngier, Zygmunt (1995). A functional theory of complementizers. In: *Modality in Grammar and Discourse.* Eds. Joan Bybee and Suzanne Fleischman. Amsterdam, Philadelphia: John Benjamins. 473–502.

Franklin, Karl J. (1971). *A Grammar of Kewa, New Guinea.* Canberra: Australian National University.

Fuß, Eric and Günther Grewendorf (2011). Zur Syntax von w-Elementen in freien Relativsätzen. Project sketch, available at <http://www.fg1783.uni-frankfurt.de/ projects.htm> (as at July 2012).

Gahl, Susanne and Alan C.L. Yu (2006). Introduction to the special issue on exemplar-based models in linguistics. *The Linguistic Review* 23: 213–216.

Galucio, Ana Vilacy (2001). The Morphosyntax of Mekens (Tupi). PhD dissertation, The University of Chicago.

Galucio, Ana Vilacy (2002). Word order and constituent structure in Mekens. In: *Revista da ABRALIN* 1.2: 51–73.

Gamerschlag, Thomas (2007). Semantic and structural aspects of complement control in Korean. In: *Studies in Complement Control.* Ed. Barbara Stiebels. *ZAS Papers in Linguistics 47.* Berlin: ZAS. 81–123.

Gast, Volker (2007). From phylogenetic diversity to structural homogeneity: On right-branching constituent order in Mesoamerica. *SKY Journal of Linguistics* 20: 171–202.

Genetti, Carol (2006). Complement clause types and complementation strategy in Dolakha Newar. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 137–158.

Genetti, Carol (2007). *A Grammar of Dolakha Newar.* Berlin, New York: Mouton de Gruyter.

Genetti, Carol, Ellen Bartee, A. R. Coupe, Kristine Hildebrandt, and You-Jing Lin (2008). Syntactic aspects of nominalization in five Tibeto-Burman languages of the Himalayan area. *Linguistics of the Tibeto-Burman Area* 31.2. 97–144.

Gerdts, Donna B. and Thomas E. Hukari (2007). A-subjects and control in Halkomelem. *Proceedings of the 7ᵗʰ International HPSG Conference.* Eds. Dan Flickinger and Andreas Kathol. Stanford: CSLI Publications. 100–123.

Gerner, Matthias (2012). The typology of nominalization. *Language and Linguistics* 13.4: 804–844.

Givón, Talmy (1979). *On Understanding Grammar.* New York: Academic Press.

Givón, Talmy (1980). The binding hierarchy and the typology of complements. *Studies in Language* 4: 333–377.

Givón, Talmy (1983). Topic continuity in discourse: An introduction. In: *Topic Continuity in Discourse: Quantitative Cross-Linguistic Studies.* Ed. Talmy Givón. Amsterdam, Philadelphia: John Benjamins. 1–42.

Givón, Talmy (1991). The evolution of dependent clause morpho-syntax in Biblival Hebrew. In: *Approaches to Grammaticalization. Vol. II.* Eds. Elizabeth Traugott and Bernd Heine. Amsterdam, Philadelphia: John Benjamins. 257–310.

Givón, Talmy (2001). *Syntax: An Introduction. Vol. II.* Amsterdam, Philadelphia: John Benjamins.

Givón, Talmy (2009). *The Genesis of Syntactic Complexity.* Amsterdam, Philadelphia: John Benjamins.

Glinert, Lewis (1989). *The Grammar of Modern Hebrew.* Cambridge: Cambridge University Press.

Göksel, Aslı and Celia Kerslake (2005). *Turkish: A Comprehensive Grammar.* London, New York: Routledge.

Goldberg, Adele E. (1995). *Constructions: A Construction Grammar Approach to Argument Structure.* Chicago, London: University of Chicago Press.

Goldberg, Adele E. (2006). *Constructions at Work: The Nature of Generalization in Language.* Oxford: Oxford University Press.

Gordon, Raymond G. Jr., ed. (2005). *Ethnologue: Languages of the World.* 15th ed. Dallas: SIL International. Available online at <http://www.ethnologue.com> (as at February 2008).

Goudswaard, Nelleke E. (2005). *The Begak (Ida'an) Language of Sabah.* Utrecht: LOT Publications.

Greenberg, Joseph H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In: *Universals of Language.* Ed. Joseph H. Greenberg. Cambridge, MA: MIT Press. 73–113.

Grenoble, Lenore A. (2012). From clause to discourse: The structure of Evenki narrative. In: *Recent Advances in Tungusic Linguistics.* Eds. Lindsay J. Whaley and Andrej Malchukov. Wiesbaden: Harrassowitz. 257–277.

Gries, Stefan Th. (2008). *Statistik für Sprachwissenschaftler.* Göttingen: Vandenhoeck & Ruprecht.

Gries, Stefan Th. (2012). Behavioral profiles: A fine-grained and quantitative approach in corpus-based lexical semantics. In: *Methodological and Analytic Frontiers in Lexical Research.* Eds. Gonia Jarema, Gary Libben and Chris Westbury. Amsterdam, Philadelphia: John Benjamins. 57–80.

Gries, Stefan Th. (2013). *Statistics for Linguistics with R: A Practical Introduction.* 2nd ed. Berlin, Boston: de Gruyter Mouton.

Grimshaw, Jane B. (1977). English *wh*-Constructions and the Theory of Grammar. PhD dissertation, University of Amherst.

Grinevald, Colette (no date). A grammar of Rama. Unpublished draft manuscript, Université de Lyon II.

Grosu, Alexander and Sandra A. Thompson (1977). Constraints on the distribution of NP clauses. *Language* 53: 104–151.

Guirardello, Raquel (1999). A Reference Grammar of Trumai. PhD dissertation, Rice University.

Güldemann, Tom (2003). Logophoricity in Africa: An attempt to explain and evaluate the significance of its modern distribution. *Sprachtypologie und Universalienforschung* 56.4: 366–387.

Güldemann, Tom (2008). *Quotative Indexes in African Languages: A Synchronic and Diachronic Survey.* Berlin, New York: Mouton de Gruyter.

Güldemann, Tom and Manfred von Roncador (eds.) (2002). *Reported Discourse: A Meeting Ground for Different Linguistic Domains.* Amsterdam, Philadelphia: John Benjamins.

Gundel, Jeanette K. (1988). *The Role of Topic and Comment in Linguistic Theory.* New York: Garland Publishing Co.

Günthner, Susanne (2009). Extraposition mit *es* im gesprochenen Deutsch. *Zeitschrift für Germanistische Linguistik* 37: 1–14.

Haig, Geoffrey (2001). Linguistic diffusion in modern East Anatolia: From top to bottom. In: *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 195–224.

Haiman, John (ed.) (1985). *Iconicity in Syntax. Proceedings of a Symposium on Iconicity in Syntax. Stanford, June 24–26, 1983.* Amsterdam, Philadelphia: John Benjamins.

Haiman, John (1989). Alienation in grammar. *Studies in Language* 13.1: 129–170.

Haiman, John and Sandra A. Thompson (1984). "Subordination" in Universal Grammar. *Berkeley Linguistics Society* 10: 510–523.

Hale, Kenneth L. (1975). Gaps in grammar and culture. In: *Linguistics and Anthropology. In Honor of C.F. Voegelin.* Eds. M. Dale Kinkade, Kenneth L. Hale and Oswald Werner. Lisse: The Peter de Ridder Press. 295-315.

Halliday, M.A.K. (1994). *Introduction to Functional Grammar.* 2nd ed. London: Arnold.

Harms, Phillip Lee. (1994). *Epena Pedee Syntax.* Summer Institute of Linguistics and University of Texas at Arlington Press.

Harriehausen, Bettina (1990). *Hmong Njua.* Tübingen: Niemeyer.

Harris, Alice C. (1981). *Georgian Syntax: A Study in Relational Grammar*. Cambridge: Cambridge University Press.

Harris, Alice C. and Lyle Campbell (1995). *Historical Syntax in Cross-Linguistic Perspective.* Cambridge: Cambridge University Press.

Hartmann, Iren, Martin Haspelmath and Bradley Taylor (eds.) (2013). *Valency Patterns Leipzig*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available online at <http://valpal.info/> (as at January 2014).

Haspelmath, Martin (1989). From purposive to infinitive – a universal path of grammaticization. *Folia Linguistica Historica* 10.1-2: 287–310.

Haspelmath, Martin. (1993). *A Grammar of Lezgian*. Berlin, New York: Mouton de Gruyter.

Haspelmath, Martin (1995). The converb as a cross-linguistically valid category. In: *Converbs in Cross-Linguistic Perspective.* Eds. Martin Haspelmath and Ekkehard König. Berlin, New York: Mouton de Gruyter. 1–55.

Haspelmath, Martin (2001). Non-canonical marking of core arguments in European languages. In: *Non-Canonical Marking of Subjects and Objects.* Eds. Alexandra Y. Aikhenvald, R.M.W. Dixon and Masayuki Onishi. Amsterdam, Philadelphia: John Benjamins. 53–83.

Haspelmath, Martin (2008). Creating economical morphosyntactic patterns in language change. In: *Language Universals and Language Change.* Ed. Jeff Good. Oxford: Oxford University Press. 185–214.

Haspelmath, Martin (2010). Comparative concepts and descriptive categories in cross-linguistic studies. *Language* 86.3: 663–687.

Haspelmath, Martin (2011a). On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15: 535–567.

Haspelmath, Martin (2011b). 'Want' complement subjects. In: *The World Atlas of Language Structures Online.* Eds. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, chapter 124. Available online at <http://wals.info/chapter/124> (as at December 2011).

Haspelmath, Martin (2013). Argument indexing: A conceptual framework for the syntax of bound person forms. In: *Languages across Boundaries: Studies in Memory of Anna Siewierska.* Eds. Dik Bakker and Martin Haspelmath. Berlin, Boston: De Gruyter Mouton. 197–226.

Haspelmath, Martin (to appear). On the cross-linguistic distribution of same-subject and different-subject 'want' complements: Economic vs. iconic motivation. *SKY Journal of Linguistics.* Available online at <http://www.eva.mpg.de/lingua/staff/haspelmath /publications_haspelmath.php> (as at November 2012).

Haspelmath, Martin, Matthew S. Dryer, David Gil and Bernard Comrie (eds.) (2005). *The World Atlas of Language Structures*. Oxford: Oxford University Press.

Hawkins, John A. (1986). *A Comparative Typology of English and German: Unifying the Contrasts.* London and Sydney: Croom Helm.

Hawkins, John A. (1994). *A Performance Theory of Word Order and Constituency.* Cambridge: Cambridge University Press.

Hawkins, John A. (2004). *Efficiency and Complexity in Grammars.* Oxford: Oxford University Press.

Hawkins, John A. (2007). Processing typology and why psychologists need to know about it. *New Ideas in Psychology* 25: 87–107.

Hayami-Allen, Rika (2001). A Descriptive Study of the Language of Ternate, the northern Moluccas, Indonesia. PhD dissertation, University of Pittsburgh.

Hayes, Bruce P. (2000). Gradient well-formedness in Optimality Theory. In: *Optimality Theory: Phonology, Syntax, and Acquisition.* Eds. Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer. Oxford: Oxford University Press. 88–120.

Heath, Jeffrey (1999). *A Grammar of Koyra Chiini.* Berlin, New York: Mouton de Gruyter.

Heath, Jeffrey (2005). *Tamashek: The Tuareg of Mali.* Berlin, New York: Mouton de Gruyter.

Heath, Jeffrey (2008). *A Grammar of Jamsay.* Berlin, New York: Mouton de Gruyter.

Heath, Jeffrey (2010). Typology of clausal boundary marking devices. *Linguistic Typology* 14.1: 127–151.

Heine, Bernd (1993). *Auxiliaries: Cognitive Forces and Grammaticalization.* Oxford: Oxford University Press.

Heine, Bernd (2009a). Africa as a linguistic area. In: *Concise Encyclopedia of Languages of the World.* Eds. Keith Brown and Sarah Ogilvie. Amsterdam etc.: Elsevier. 3–7.

Heine, Bernd (2009b). From nominal to clausal morphosyntax: Complexity via expansion. In: *Syntactic Complexity: Diachrony, Acquisition, Neuro-Cognition, Evolution.* Eds. Talmy Givón and Masayoshi Shibatani. Amsterdam, Philadelphia: John Benjamins. 23–51.

Heine, Bernd and Tania Kuteva (2002). *World Lexicon of Grammaticalization.* Cambridge: Cambridge University Press.

Heine, Bernd and Tania Kuteva (2007). *The Genesis of Grammar: A Reconstruction.* Oxford: Oxford University Press.

Heine, Bernd and Derek Nurse (eds.) (2008). *A Linguistic Geography of Africa.* Cambridge: Cambridge University Press.

Hellwig, Bigit (2006). Complement clause type and complementation strategies in Goemai. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 204–223.

Hendery, Rachel (2012). *Relative Clauses in Time and Space.* Amsterdam, Philadelphia: John Benjamins.

Hengeveld, Kees (1989). Layers and operators in Functional Grammar. *Journal of Linguistics* 25: 127–157.

Hennesy, Pelin and Talmy Givón (2001). Note on the grammar of Turkish nominalizations. In: *Complex Sentences in Grammar and Discourse. Essays in Honor of Sandra A. Thompson.* Eds. Joan Bybee and Michael Noonan. Amsterdam, Philadelphia: John Benjamins. 125–144.

Hewitt, Brian G. (1987). *The Typology of Subordination in Georgian and Abkhaz.* Berlin, New York: Mouton de Gruyter.

Hewitt, Brian G. (1995). *Georgian: a Structural Reference Grammar.* Amsterdam, Philadelphia: John Benjamins.

Hinds, John (1986). *Japanese.* London: Routledge.

Höftmann, Hildegard (1993). *Grammatik des Fon.* Leipzig: Langenscheidt, Akademie Verlag.

Holes, Clive (1990). *Gulf Arabic.* London: Routledge.

Holland, John H. (2006). Studying complex adaptive systems. *Journal of Systemic Science and Complexity* 19: 1–8.

Hollenbach, Elena E. (1995). Cuatro morfemas funcionales en las lenguas mixtecanas. In: *Vitalidad y Influencia de las Lenguas Indígenas en Latinoamérica: Il Coloquio Mauricio Swadesh.* Eds. Ramón Arzápalo Marín and Yolanda Lastra. UNAM: Instituto de Investigaciones Antropológicas. 284–293.

Holton, Gary (2003). *Tobelo.* München: Lincom Europa.

Hopper, Paul J. (1987). Emergent grammar. *Berkeley Linguistics Society* 13: 139–157.

Hopper, Paul J. (1991). On some principles of grammaticization. In: *Approaches to Grammaticalization. Vol. I: Focus on Theoretical and Methodological Issues.* Eds. Elizabeth C. Traugott and Bernd Heine. Amsterdam/Philadelphia: John Benjamins. 17–35.

Hopper Paul J. (2004). The openness of grammatical constructions. *Chicago Linguistics Society* 40: 153–175.

Hopper, Paul J. and Sandra A. Thompson (1980). Transitivity in grammar and discourse. *Language* 56: 251–299.

Hopper, Paul J. and Elizabeth C. Traugott (2003). *Grammaticalization.* 2nd ed. Cambridge: Cambridge University Press.

Horie, Kaoru (1993). A Cross-Linguistic Study of Perception and Cognition Verb Complements: A Cognitive Perspective. PhD dissertation, University of Southern California, Los Angeles.

Horie, Kaoru (2001). Complement clauses. In: *Handbook of Language Typology and Language Universals.* Eds. Martin Haspelmath, Ekkehard König, Wulf Oesterreicher, and Wolfgang Raible. Berlin, New York: Walter de Gruyter. 979–993.

Howe, Mark L. and Marc D. Lewis (2005). The importance of dynamic systems approaches for understanding development. *Developmental Review* 25: 247–251.

Hualde, José Ignacio and Jon Ortiz de Urbina, eds. (2003). *A Grammar of Basque.* Berlin: Mouton de Gruyter.

Huang, Shuanfan (2003). Doubts about complementation: A functionalist analysis. *Language and Linguistics* 4.2: 429–455.

Huddleston, Rodney and Geoffrey K. Pullum (2002). *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.

Huson, Daniel H. and David Bryant (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23.2: 254–267.

Huttar, George L. and Mary L. Huttar (1994). *Ndyuka.* London: Routledge.

Ikoro, Suanu M. (1996). *The Kana Language.* Leiden: Research School CNWS.

Inaba, Jiro (2007). *Die Syntax der Satzkomplementierung: Zur Struktur des Nachfeldes im Deutschen.* Berlin: Akademie Verlag.

Jacobson, Steven A. (1995). *A Practical Grammar of the Central Alaskan Yup'ik Eskimo Language.* Fairbanks: Alaska Native Language Center.

Jaeger, T. Florian (2010). Redundancy and reduction: Speakers manage information density. *Cognitive Psychology* 61.1: 23–62.

Jäger, Andreas (2007). Coding strategies of indirect reported speech in typological perspective. *University of Queensland Working Papers in Linguistics* 1. Available online at <http://www.library.uq.edu.au/ojs/index.php/uqwpl/article/view/16/17> (as at November 2010).

Jaggar, Philip J. (2001). *Hausa.* Amsterdam, Philadelphia: John Benjamins.

Janssen, Dirk, Balthasar Bickel and Fernando Zúñiga (2006). Randomization tests in language typology. *Linguistic Typology* 10: 419–440.

Jarkey, Nerida (2006). Complement clause types and complementation strategy in White Hmong. In: *Complementation: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 115–136.

Jayaseelan, K.A. (1999). *Parametric Studies in Malayalam Syntax.* Mumbai: Allied Publishers.

Jayaseelan, K.A. (2004). The possessor-experiencer dative in Malayalam. In: *Non-Nominative Subjects. Vol. I.* Eds. Peri Bhaskararao and Karumuri Venkata Subbarao. Amsterdam, Philadelphia: John Benjamins. 227–244.

Jelinek, Eloise (1984). Empty categories and non-configurational languages. *Natural Language and Linguistic Theory* 2: 39–76.

Johanson, Lars (2011). Selection of subjunctors in Turkic complement clauses. Paper presented at the workshop on *Complementation in Eurasian Languages*, University of Mainz, November 2011.

Johnson, Audrey F. (1988). A syntactic sketch of Jamiltepec Mixtec. In: *Studies in the Syntax of Mixtecan Languages. Vol. I.* Eds. C. Henry Bradley and Barbara E. Hollenbach. The Summer Institute of Linguistics and The University of Texas at Arlington Publications. 11–150.

Johnson, Heidi A. (2000). A Grammar of San Miguel Chimalapa Zoque. PhD dissertation, University of Texas at Austin.

Jones, Wendell and Paula Jones (1991). *Barasano Syntax.* Arlington: The Summer Institute of Linguistics and The University of Texas at Arlington.

Joseph, Brian D. (1983). *The Synchrony and Diachrony of the Balkan Infinitive.* Cambridge: Cambridge University Press.

Josephs, Lewis (1976). Complementation. In: *Syntax and Semantics: Japanese Generative Grammar. Vol. 5.* Ed. Masayoshi Shibatani. New York: Academic Press. 307–369.

Justeson, John S. and Laurence D. Stephens (1990). Explanations for word order universals: A log-linear analysis. In: *Proceedings of the XIV International Congress of Linguists. Vol. 3.* Eds. Werne Bahner, Joachim Schildt and Dieter Viehweger. Berlin: Akademie Verlag. 2372–2376.

Kaiser, Stefan, Yasuko Ichikawa, Noriko Kobayashi and Hilofumi Yamamoto (2013). *Japanese: A Comprehensive Grammar.* 2nd ed. Abingdon, New York: Routledge.

Kaltenböck, Gunther (2004). It-*Extraposition and Non-Extraposition in English.* Wien: Braumüller.

Kaufman, Leonard and Peter J. Rousseeuw (1990). *Finding Groups in Data: An Introduction to Cluster Analysis.* New York: Wiley.

Kazenin, Konstantin I. (1994). Split Syntactic Ergativity: Toward an Implicational Hierarchy. *Sprachtypologie und Universalienforschung* 47: 78–98.

Keenan, Edward L. (1978). The syntax of subject-final languages. In: *Syntactic Typology. Studies in the Phenomenology of Language.* Ed. Winfred P. Lehmann. Sussex: The Harvester Press. 267–327.

Keenan, Edward L. and Bernard Comrie (1977). Noun phrase accessibility and universal grammar. *Linguistic Inquiry* 8: 63–99.

Kemmer, Suzanne and Michael Barlow (2000). Introduction: A usage-based conception of language. In: *Usage-Based Models of Language.* Eds. Michael Barlow and Suzanne Kemmer. Stanford, CA: Center for the Study of Language and Information. vii–xxviii.

Kenesei, István (1994). Subordinate clauses. In: *The Syntactic Structure of Hungarian. Syntax and Semantics 27.* Eds. Ferenc Kiefer and Katalin É. Kiss. San Diego: Academic Press. 275–354.

Kenesei, István, Robert M. Vago and Anna Fenyvesi (1998). *Hungarian.* London, New York: Routledge.

Keresztes, László (1999). *Praktische ungarische Grammatik.* 3rd ed. Debrecen: Debreceni Nyári Egyetem.

Khanina, Olesya (2008). How universal is *wanting*? *Studies in Language* 32.4: 818–865.

Khanina, Olesya (2009). The symbiosis of descriptive linguistics and typology: A case study of desideratives. In: *New Challenges in Typology: Transcending the Borders and Redefining the Distinctions.* Eds. Patience Epps and Alexandre Arkhipov. Berlin, New York: Mouton de Gruyter. 199–220.

Kilian-Hatz, Christa (1999). *Folktales of the Kxoe in the West Caprivi.* Köln: Rüdiger Köppe.

Kilian-Hatz, Christa (2008). *A Grammar of Modern Khwe.* Köln: Rüdiger Köppe.

Kim, Bo Ra (2010). Non-Finite Complements in Russian, Serbian/Croatian, and Macedonian. PhD dissertation, University of Indiana.

Kim, Jong-Bok (2014). English copy raising constructions: Argument realization and characterization condition. *Linguistics* 52.1: 167–203.

Kiparsky, Paul and Carol Kiparsky (1970). Fact. In: *Progress in Linguistics.* Eds. Manfred Bierwisch and Karl Erich Heidolph. The Hague: Mouton. 143–173.

Kirk, John W.C. (1905). *A Grammar of the Somali Language.* Cambridge: Cambridge University Press.

Kirsner, Robert S. and Sandra A. Thompson (1976). The role of pragmatic inference in semantics: A study of sensory verb complements in English. *Glossa* 10: 200–240.

Kiss, Katalin E. (2002). *The Syntax of Hungarian.* Cambridge: Cambridge University Press.

Kiyosawa, Kaoru (2006). Applicatives in Salish Languages. PhD dissertation, Simon Fraser University.

Klaiman, M.H. (1976). A Functional View of some Syntactic Movement Typologies. MA thesis, University of Chicago.

Klamer, Marian (2000). How report verbs become quote markers and complementisers. *Lingua* 110: 69–98.

Klamer, Marian (2002). 'Report' constructions in Kambera (Austronesian). In: *Reported Discourse.* Eds. Tom Güldemann and Manfred von Roncador. Amsterdam, Philadelphia: John Benjamins. 323–340.

Knudson, L. (1980). *Zoque de Chimalapa.* San Ángel: El Colegio de México.

König, Ekkehard and Volker Gast (2012). *Understanding English-German Contrasts.* 3rd ed. Berlin: Erich Schmidt Verlag.

Koptjevskaja-Tamm, Maria (1993). *Nominalizations.* London: Routledge.

Kordić, Snjezana (1997). *Serbo-Croatian.* München: Lincom Europa.

Kordić, Snjezana (1999). *Der Relativsatz im Serbokroatischen.* München: Lincom Europa.

Kornfilt, Jaklin (1997). *Turkish.* London: Routledge.

Kremers, Joost M. (2003). *The Arabic Noun Phrase: A Minimalist Approach.* Utrecht: LOT Publications.

Krishnamurti, Bhadriraju (2003). *The Dravidian Languages.* Cambridge: Cambridge University Press.

Kroeber, Paul D. (1999). *The Salish Language Family: Reconstructing Syntax.* The University of Nebraska Press.

Kroeger, Paul R. (2005). *Analyzing Grammar: An Introduction.* Cambridge: Cambridge University Press.

Kruskal, Joseph B. and Myron Wish (1978). *Multidimensional Scaling.* Newbury Park etc.: Sage Publications.

Kruspe, Nicole (2004). *A Grammar of Semelai.* Cambridge: Cambridge University Press.

Kung, Susan S. (2007). A Descriptive Grammar of Huehuetla Tepehua. PhD dissertation, University of Texas at Austin.

Kuno, Susumu (1972). Evidence for subject raising in Japanese. *Papers in Japanese Linguistics* 1.1: 24–51.

Kuno, Susumu (1973). *The Structure of the Japanese Language*. Cambridge, MA: MIT Press.

Kuno, Susumo (1974). The position of relative clauses and conjunctions. *Linguistic Inquiry* 5: 117–136.

Kutscher, Silvia (2009). *Kausalität und Argumentrealisierung: Zur Konstruktionsvarianz bei Psychverben am Beispiel europäischer Sprachen*. Tübingen: Niemeyer.

Lamberti, Marcello and Roberto Sottile (1997). *The Wolaytta Language*. Köln: Rüdiger Köppe Verlag.

Lambton, Ann K.S. (1974). *Persian Grammar*. Cambridge: Cambridge University Press.

Landau, Idan (2000). *Elements of Control: Structure and Meaning in Infinitival Constructions*. Dordrecht: Kluwer.

Landau, Idan (2009). This construction looks like a copy is optional. *Linguistic Inquiry* 40.2: 343–346.

Landau, Idan (2011). Predication vs. aboutness in copy raising. *Natural Language and Linguistic Theory* 29: 779–813.

Langacker, Ronald W. (1987). *Foundations of Cognitive Grammar. Vol. I: Theoretical Prerequisites*. Stanford: Stanford University Press.

Langacker, Ronald W. (1991). *Foundations of Cognitive Grammar. Vol. II: Descriptive Application*. Stanford: Stanford University Press.

Langacker, Ronald W. (1995). Raising and transparency. *Language* 71.1: 1–62.

Langacker, Ronald W. (2000). A dynamic usage-based model. In: *Usage-Based Models of Language*. Eds. Michael Barlow and Suzanne Kemmer. Stanford, CA: Center for the Study of Language and Information. 1–63.

Langacker, Ronald W. (2009). *Investigations in Cognitive Grammar*. Berlin, New York: Mouton de Gruyter.

LaPolla, Randy J. (1990). Grammatical Relations in Chinese: Synchronic and Diachronic Considerations. PhD dissertation, University of California at Berkeley.

LaPolla, Randy J. (1993). Arguments against 'subject' and 'object' as viable concepts in Chinese. *Bulletin of the Institute of History and Philology* 63.4: 759–813.

LaPolla, Randy J. (2001). The role of migration and language contact in the development of the Sino-Tibetan language family. In: *Areal Diffusion and Genetic Inheritance: Problems in Comparative Linguistics*. Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 225–254.

Lazard, Gilbert (2002). Transitivity revisited as an example of a more strict approach in typological research. *Folia Linguistica* 36: 141–190.

Lee, Sun-Hee (1994). A Lexical Analysis of Selected Unbounded Dependency Constructions in Korean. PhD dissertation, The Ohio State University.

Lefebvre, Claire (ed.) (1991). *Serial Verbs: Grammatical, Comparative and Cognitive Approaches*. Amsterdam, Philadelphia: John Benjamins.

Lefebvre, Claire and Anne-Marie Brousseau (2002). *A Grammar of Fongbe*. Berlin, New York: Mouton de Gruyter.

Lehmann, Christian (1984). *Der Relativsatz: Typologie seiner Strukturen. Theorie seiner Funktionen. Kompendium seiner Grammatik*. Tübingen: Gunter Narr.

Lehmann, Christian (1986). On the typology of relative clauses. *Linguistics* 24: 663–680.

Lehmann, Christian (1988). Towards a typology of clause linkage. In: *Clause Combining in Grammar and Discourse*. Eds. John Haiman and Sandra A. Thompson. Amsterdam, Philadelphia: John Benjamins. 181–226.

Lehmann, Christian (2002). *Thoughts on Grammaticalization*. 2nd ed. Erfurt: Arbeitspapiere des Seminars für Sprachwissenschaft der Universität Erfurt.

Levy, Roger and T. Florian Jaeger (2007). Speakers optimize information density through syntactic reduction. In: *Proceedings of the Twentieth Annual Conference on Neural Information Processing Systems*, Vancouver, British Columbia, Canada, December 4-7, 2006. Eds. Bernhard Schölkopf, John C. Platt and Thomas Hoffman. Cambridge, MA.: MIT Press. 849–856.

Lewis, Geoffrey L. (1967). *Turkish Grammar*. Oxford: Oxford University Press.

Li, Charles N. and Sandra A. Thompson (1976). Subject and topic: A new typology of language. In: *Subject and Topic*. Ed. Charles N. Li. New York etc.: Academic Press. 457–489.

Li, Charles N. and Sandra A. Thompson (1981). *Mandarin Chinese: A Functional Reference Grammar.* Berkeley etc.: University of California Press.

Lichtenberk, Frantisek (1991). On the gradualness of grammaticalization. In: *Approaches to Grammaticalization. Vol. I.* Eds. Elizabeth Closs Traugott and Bernd Heine. Amsterdam, Philadelphia: John Benjamins. 37–80.

Lichtenberk, Frantisek (2008). *A Grammar of Toqabaqita.* Berlin, New York: Mouton de Gruyter.

Linn, Mary Sarah (2001). A Grammar of Euchee (Yuchi). PhD dissertation, University of Kansas.

Longacre, Robert E. (1968). *Philippine Languages: Discourse, Paragraph and Sentence Structure. Vol. I.* Santa Ana, CA: Summer Institute of Linguistics.

Longacre, Robert E. (1976). *An Anatomy of Speech Notions.* Lisse: Peter de Ridder.

Lord, Carol (1993). *Historical Change in Serial Verb Constructions.* Amsterdam, Philadelphia: John Benjamins.

Los, Bettelou (2005). *The Rise of the To-Infinitive.* Oxford: Oxford University Press.

Lotfi, Ahmad R. (2008). Causative constructions in Modern Persian. *California Linguistic Notes* XXXIII.2. Available online at <http://hss.fullerton.edu/linguistics/cln/ Sp%2008%20pdf/Lofti-Causative%20Constructions.pdf> (as at July 2012).

Lühr, Rosemarie (2008). Competitive Indo-European syntax. In: *Principles of Syntactic Reconstruction.* Eds. Gisella Ferraresi and Maria Goldbach. Amsterdam, Philadelphia: John Benjamins. 121–160.

Lyons, John (1968). *Introduction to Theoretical Linguistics.* Cambridge: Cambridge University Press.

Macaulay, Monica (1996). *A Grammar of Chalcatongo Mixtec.* Berkeley etc.: University of California Press.

Macaulay, Monica (2005). The syntax of Chalcatongo Mixtec: Preverbal and postverbal. *Linguistics Today* 73: 341–366.

MacWhinney Brian and Csaba Pléh (1988). The processing of restrictive relative clauses in Hungarian. *Cognition* 29.2: 95–144.

Mace, John (2003). *Persian Grammar.* London: Routledge.

Mahootian, Shahrzad (1997). *Persian.* London and New York: Routledge.

Mair, Christian (1990). *Infinitival Complement Clauses in English: A Study of Syntax in Discourse.* Cambridge: Cambridge University Press.

Malchukov, Andrej (2004). *Nominalization/Verbalization: Constraining a Typology of Trans-categorial Operations.* München: Lincom Europa.

Malchukov, Andrej (2008). Split intransitives, experiencer objects and 'transimpersonal' constructions: (re)-establishing the connection. In: *The Typology of Semantic Alignment.* Eds. Mark Donohue and Søren Wichmann. Oxford: Oxford University Press. 76–100.

Manning, Christopher D. and Hinrich Schütze (1999). *Foundations of Statistical Natural Language Processing.* Cambridge, MA: MIT Press.

Maslova, Elena (2003). *A Grammar of Kolyma Yukaghir.* Berlin, New York: Mouton de Gruyter.

Matthews, Peter H. (1997). *Oxford Concise Dictionary of Linguistics.* Oxford: Oxford University Press.

Matthiesen, Christian and Sandra A. Thompson (1988). The structure of discourse and "subordination". In: *Clause Combining in Grammar and Discourse.* Eds. John Haiman and Sandra A. Thompson. Amsterdam, Philadelphia: John Benjamins. 275–333.

Maynard, Senko K. (1986). The particle *-o* and content-oriented indirect speech in Japanese written discourse. In: *Direct and Indirect Speech.* Ed. Florian Coulmas. Berlin, New York: Mouton de Gruyter. 179–200.

McConvell, Patrick (2006). Grammaticalization of demonstratives as subordinate complementizers in Ngumpin-Yapa. *Australian Journal of Linguistics* 26.1: 107–137.

McGregor, William (1988). Mood and subordination in Kuniyanti. In: *Complex Sentence Constructions in Australian Languages.* Ed. Peter Austin. Amsterdam, Philadelphia: John Benjamins. 37–67.

McGregor, William (1990). *A Functional Grammar of Gooniyandi.* Amsterdam, Philadelphia: John Benjamins.

McGregor, William B. (1994). The grammar of reported speech and thought in Gooniyandi. *Australian Journal of Linguistics* 14.1: 63–92.

McGregor, William B. (2013). Optionality in grammar and language use. *Linguistics* 51.6: 1147–1204.

Menn, Lise and Brian MacWhinney (1984). The repeated morph constraint: Toward an explanation. *Language* 60.3: 519–541.

Merlan, Francesca (1981). Some functional relations among subordination, mood, aspect and focus in Australian languages. *Australian Journal of Linguistics* 1: 175–210.

Merlan, Francesca (1982). *Mangarayi.* Amsterdam: North-Holland.

Miller, Amy W. (2001). *A Grammar of Jamul Tiipay.* Berlin, New York: Mouton de Gruyter.

Miller, John M. (1989). Sentential Complementation in Huaraz Quechua: A Typological-Functional Study. MA thesis, University of Texas at Arlington.

Miller, Philip H. (2001). Discourse constraints on (non)extraposition from subject in English. *Linguistics* 39.4: 683–701.

Mollica, Fabio (2010). *Korrelate im Deutschen und im Italienischen.* Frankfurt am Main etc.: Peter Lang.

Munro, Pamela (1982). On the transitivity of 'say' verbs. In: *Studies in Transitivity.* Eds. Paul J. Hopper and Sandra A. Thompson. New York: Academic Press. 301–318.

Muysken, Pieter (1994). Nominalizations. In: *The Encyclopedia of Language and Linguistics.* Eds. Ronald E. Asher and James M.Y. Simpson. Oxford: Pergamon Press. 2811–2815.

Næss, Åshild (2007). *Prototypical Transitivity.* Amsterdam, Philadelphia: John Benjamins.

Nedjalkov; Igor' V. (1997). *Evenki.* London, New York: Routledge.

Nedjalkov, Igor' V. (1998) Converbs in the languages of Europe. In: *Adverbial Constructions in the Languages of Europe.* Eds. Johan van der Auwera and Dónall P. Ó Baoill. Berlin, New York: Mouton de Gruyter. 421–455.

Neukom, Lukas (2001). *Santali.* München: Lincom Europa.

Newman, Paul (2000). *The Hausa Language. An Encyclopedic Reference Grammar.* New Haven, London: Yale University Press.

Newmeyer, Frederick J. (1975). *English Aspectual Verbs.* The Hague: Mouton.

Nguyên, Đình-Hoà (1988). An outline of Vietnamese. *The Vietnam Forum* 11: 1–20.

Nguyên, Đình-Hoà (1996). Vietnamese verbs. *Mon-Khmer Studies* 25: 141–159.

Nguyên, Đình-Hoà (1997). *Vietnamese.* Amsterdam, Philadelphia: John Benjamins.

Nguyên, Thanh Húng (1979). *Einführung in die vietnamesische Sprache.* Frankfurt: Haag und Herchen.

Nichols, Johanna (1992). *Linguistic Diversity in Space and Time.* Chicago: University of Chicago Press.

Nicklas, Thurston Dale (1974). Elements of Choctaw Grammar. PhD dissertation, University of Michigan.

Noonan, Michael (1992). *A Grammar of Lango.* Berlin, New York: Mouton de Gruyter.

Noonan, Michael (2007) [1985]. Complementation. In: *Language Typology and Syntactic Description. Vol. II: Complex Constructions.* Ed. Timothy Shopen. Cambridge: Cambridge University Press. 52–150.

Nordlinger, Rachel (1998). *A Grammar of Wambaya, Northern Territory (Australia).* Canberra: Australian National University.

Nordlinger, Rachel (2002). Non-finite subordinate verbs in Australian aboriginal languages: Are nominalised verbs really nominalised? In: *Proceedings of the 2001 Conference of the Australian Linguistic Society.* Available online at <http://www.als.asn.au/proceedings/als2001/nordlinger.pdf> (as at August 2011).

Nordlinger, Rachel (2006). Spearing the emu drinking: Subordination and the adjoined relative clause in Wambaya. *Australian Journal of Linguistics* 26.1: 5–29.

O'Dowd, Elizabeth (1992). The syntactic metaphor of subordination: A typological study. *Lingua* 86: 47–80.

Öhl, Peter and Agnes Korn (2006). Performanzbasierte und parametrische Wandel in der linken Satzperipherie des Persischen: Der Subordinationsmarker *ke* und die Interrogativpartikel *āyā. Die Sprache* 46.2: 137–202.

Öhl, Peter and Ahmad R. Lotfi (2007). Nominalised CPs in Persian: a parametric account. Paper presented at the 2nd International Conference on Iranian Linguistics, Kiel 2007. Available online at <http://oehl.gesus-info.de/> (as at June 2012).

Olawsky, Knut J. (2002). *Urarina Texts.* München: Lincom Europa.

Olawsky, Knut J. (2006). *A Grammar of Urarina.* Berlin, New York: Mouton de Gruyter.

Olsen, Susan (1981). *Problems of* seem/scheinen *Constructions and their Implications for the Theory of Predicate Sentential Complementation.* Tübingen: Niemeyer.

Olszok, Klaus (1983). Infinitive Formen im Vorfeld. In: *Zur Wortstellungsproblematik im Deutschen.* Eds. Klaus Olszok and Edith Weuster. Tübingen: Narr. 89–169.

Onishi, Masayuki (1994). A Grammar of Motuna. PhD dissertation, Australian National University.

Onishi, Masayuki (2000). Transitivity and valency-changing derivations in Motuna. In: *Changing Valency: Case Studies in Transitivity.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Cambridge: Cambridge University Press. 115–144.

Onishi, Masayuki (2001). Non-canonically marked subjects and objects: Parameters and properties. In: *Non-Canonical Marking of Subjects and Objects.* Eds. Alexandra Y. Aikhenvald, R.M.W. Dixon and Masayuki Onishi. Amsterdam, Philadelphia: John Benjamins. 1–51.

Onishi, Masayuki (2004). Instrumental subjects in Motuna. In: *Non-Nominative Subjects. Vol. II.* Eds. Peri Bhaskararao and Karumuri Venkata Subbarao. Amsterdam, Philadelphia: John Benjamins. 83–101.

Ono, Tsuyoshi (1996). Information Flow and Grammatical Structure in Barabareño Chumash. PhD dissertation, University of California at Santa Barbara.

Palmer, Bill (1999). A Grammar of the Kokota Language, Santa Isabel, Solomon Islands. PhD dissertation, University of Sydney.

Palmer, Frank R. (2001). *Mood and Modality.* 2nd ed. Cambridge: Cambridge University Press.

Paranhos Zitterbart, Jussara (2002). *Zur korrelativen Subordination im Deutschen.* Tübingen: Niemeyer.

Park, Ki-Seong (1993). Korean Causatives in Role and Reference Grammar. MA thesis, State University of New York at Buffalo.

Park, Ki-Seong (1995). The Semantics and Pragmatics of Case Marking in Korean: A Role and Reference Grammar Account. PhD dissertation, State University of New York at Buffalo.

Payne, Doris L. (1985). Aspects of the Grammar of Yagua: A Typological Approach. PhD dissertation, UCLA.

Payne, Doris L. and Thomas E. Payne (1990). Yagua. In: *Handbook of Amazonian Languages. Vol. 2.* Eds. Desmond C. Derbyshire and Geoffrey K. Pullum. Berlin: Mouton de Gruyter. 249–474.

Payne, Thomas E. (1997). *Describing Morphosyntax.* Cambridge: Cambridge University Press.

Perkins, Revere D. (2001). Sampling procedures and statistical methods. In: *Language Typology and Language Universals: an International Handbook.* Eds. Martin Haspelmath, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible. Berlin, New York: de Gruyter. 419–434.

Perlmutter, David M. (1970). The two verbs 'begin'. In: *Readings in Transformational Grammar.* Eds. Roderick A. Jacobs and Peter S. Rosenbaum. Waltham, Mass.: Ginn & Co. 107–119.

Perlmutter, David M. and Paul M. Postal (1983). The relational succession law. In: *Studies in Relational Grammar, Vol. I.* Ed. David M. Perlmutter. Chicago and London: University of Chicago Press. 30–80.

Persson, Maria (2002). *Sentential Object Complements in Modern Standard Arabic.* Studia Orientalia Lundensia, Nova Series, 2. Stockholm: Almqvist & Wiksell.

Pesetsky, David (1995). *Zero Syntax: Experiencer and Cascades.* Cambridge, MA: MIT Press.

Philipose, Skaria C. (1984). *A Constrastive Study of Case in English and Malayalam.* Frankfurt: Peter Lang.

Pinker, Steven (1989). *Learnability and Cognition.* Cambridge, MA: MIT Press.

Postal, Paul M. (1974). *On Raising.* Cambridge, MA and London: MIT Press.

Potsdam, Eric and Jeffrey Runner (2001). Richard returns: Copy raising and its implications. In: *Papers from the 37th Regional Meeting of the Chicago Linguistic Society (CLS 37).* Eds. Mary Andronis, Chris Ball, Heidi Elston and Sylvain Neuvel. Chicago, IL: Chicago Linguistic Society. 453–468.

R Development Core Team (2010). *R: A Language and Environment for Statistical Computing.* Version 2.12.0. Vienna: R Foundation for Statistical Computing. Available online at <http://www.r-project.org> (as at November 2010).

Ransom, Evelyn R. (1986). *Complementation: Its Meanings and Forms.* Amsterdam: John Benjamins.

Rebrus, Péter and Anna Babarczy (n.d.). Hungarian Descriptive Grammar. Manuscript.

Reckendorf, Hermann (1898). *Die syntaktischen Verhältnisse des Arabischen.* Leiden: Brill.

Refsing, Kirsten (1986). *The Ainu Language: The Morphology and Syntax of the Shizunai Dialect.* Aarhus: Aarhus University Press.

Reh, Mechthild (1985). *Die Krongo-Sprache (Nìinò Mó-Dì).* Berlin: Dietrich Reimer Verlag.

Rice, Keren D. (1986). Some remarks on direct and indirect discourse in Slave (Northern Athapaskan). In: *Direct and Indirect Speech.* Ed. Florian Coulmas. Berlin, New York: Mouton de Gruyter. 47–76.

Rice, Keren (1989). *A Grammar of Slave.* Berlin, New York: Mouton de Gruyter.

Rijkhoff, Jan, Dik Bakker, Kees Hengeveld and Peter Kahrel (1993). A method of language sampling. *Studies in Language* 17: 169–203.

Rijkhoff, Jan and Dik Bakker (1998). Language sampling. *Linguistic Typology* 2: 263–314.

Roberts, John R. (1987). *Amele.* London: Croom Helm.

Roberts, John R. (2001). Impersonal constructions in Amele. In: *Non-Canonical Marking of Subjects and Objects.* Eds. Alexandra Y. Aikhenvald, R.M.W. Dixon and Masayuki Onishi. Amsterdam, Philadelphia: John Benjamins. 201–250.

Rohdenburg, Günter (2003). Cognitive complexity and *horror aequi* as factors determining the use of interrogative clause linkers in English. In: *Determinants of Grammatical Variation in English.* Eds. Günter Rohdenburg and Britta Mondorf. Berlin, New York: Mouton de Gruyter. 205–250.

Romero-Figeroa, Andrés (1997). *A Reference Grammar of Warao.* München: LINCOM Europa.

Rosenbaum, Peter (1967). *The Grammar of English Predicate Complement Constructions.* Cambridge, MA.: MIT Press.

Ruhlen, Merritt (1991). *A Guide to the World's Languages. Vol.1: Classification.* Stanford: Stanford University Press.

Rumsey, Alan (1982). *An Intra-Sentence Grammar of Ungarinjin, North-Western Australia.* Canberra: Australian National University.

Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik (1985). *A Comprehensive Grammar of the English Language.* London: Longman.

Saeed, John (1999). *Somali.* Amsterdam, Philadelphia: John Benjamins.

Sakel, Jeanette (2004). *A Grammar of Mosetén.* Berlin, New York: Mouton de Gruyter.

Saltarelli, Mario (1988). *Basque.* London, New York: Croom Helm.

Sawka, Kenneth S. (2001). Aspects of Mayogo Grammar. M.A. thesis, The University of Texas at Arlington.

Saxena, Anju (1995). Unidirectional grammaticalization: diachronic and cross-linguistic evidence. *Sprachtypologie und Universalienforschung* 48.4: 350–372.

Schegloff, Emanuel A. (2001). Discourse as an interactional achievement III: The omnirelevance of action. In: *The Handbook of Discourse Analysis.* Eds. Deborah Schiffrin, Deborah Tannen and Hedi E. Hamilton. Oxford: Blackwell. 229–249.

Schmidtke-Bode, Karsten (2009). *A Typology of Purpose Clauses.* Amsterdam, Philadelphia: John Benjamins.

Schmidtke-Bode, Karsten (2012). The performance basis of grammatical constraints on complex sentences: A preliminary survey. In: *Clause Combining in Cross-Linguistic Perspective: Data-Driven Approaches to Cross-Clausal Syntax.* Eds. Volker Gast and Holger Diessel. Berlin: Mouton de Gruyter. 415–448.

Schmidtke-Bode, Karsten and Holger Diessel (subm.). Cross-linguistic patterns in the structure, function and position of (object) complement clauses. *Linguistics.*

Schulte, Kim (2007). What causes adverbial infinitives to spread? Evidence from Romance. *Language Sciences* 29: 512–537.

Schütze, Carson T. (1996) *The Empirical Base of Linguistics: Grammaticality Judgments and Linguistic Methodology.* Chicago: University of Chicago Press.

Searle, John R. (1983). *Intentionality.* Cambridge: Cambridge University Press.

Seiler, Walter (1985). *Imonda.* Canberra: Australian National University.

Seppänen, Aimo (2002). On analysing the pronoun *it. English Studies* 5: 442–462.

Serdobolskaya, Natalia (2009). Towards the typology of raising: A functional approach. In: *New Challenges in Typology: Transcending the Borders and Redefining the Distinctions.* Eds. Patience Epps and Alexandre Arkhipov. Berlin, New York: Mouton de Gruyter. 269–294.

Serzant, Ilja A. and Leonid Kulikov (eds.) (2013). *The Diachronic Typology of Non-Canonical Subjects.* Amsterdam, Philadelphia: John Benjamins.

Shibatani, Masayoshi (1990). *The Languages of Japan.* Cambridge: Cambridge University Press.

Shibatani, Masayoshi (1991). Grammaticization of topic into subject. In: *Approaches to Grammaticalization. Vol. II.* Eds. Elizabeth Traugott and Bernd Heine. Amsterdam, Philadelphia: John Benjamins. 93–133.

Shibatani, Masayoshi (1999). Dative-subject constructions twenty-two years later. *Studies in the Linguistic Sciences* 29.2: 46–76.

Shibatani, Masayoshi (2011). Nonperiphrastic causative constructions. In: *The World Atlas of Language Structures Online.* Eds. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, chapter 111. Available online at <http://wals.info/chapter/111> (as at March 2012).

Siewierska, Anna (1993). On the interplay of factors in the determination of word order. In: *Syntax: An International Handbook. Vol. 1.* Eds. Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld and Theo Vennemann. Berlin, New York: Walter de Gruyter. 826–846.

Sierwierska, Anna (2001). On the argument status of cross-referencing forms. In: *Challenges and Developments in Functional Grammar.* Ed. Mará Jesús Perez Quintero. La Laguna, Tenerife: Universidad de La Laguna. 215–236.

Siewierska, Anna (2008). Introduction: Impersonalization from a subject-centred versus agent-centred perspective. *Transactions of the Philological Society* 106.2: 115–137.

Siewierska, Anna (2011). Alignment of verbal person marking. In: *The World Atlas of Language Structures Online.* Eds. Matthew S. Dryer and Martin Haspelmath. Munich: Max Planck Digital Library, chapter 100. Available online at <http://wals.info/chapter/100> (as at July 2012).

Silverstein, Michael (1976). Hierarchy of features and ergativity. In: *Grammatical Categories in Australian Languages.* Ed. R.M.W. Dixon. Canberra: Australian Institute of Aboriginal Studies; Atlantic Highlands, NJ: Humanities Press. 112–171.

Skopeteas, Stavros, Caroline Féry and Rusudan Asatiani (2007). Word order and intonation in Georgian. Ms., University of Potsdam.

Skribnik, Elena (2013). Mongolic clausal complementation and information source. Ms., University of Munich.

Słodowicz, Szymon (2007). Complement control in Turkish. In: *Studies in Complement Control.* Ed. Barbara Stiebels. *ZAS Papers in Linguistics 47.* Berlin: ZAS. 125–157.

Smeets, Ineke (1989). A Mapuche Grammar. PhD dissertation, University of Leiden.

Smeets, Ineke (2008). *A Grammar of Mapuche.* Berlin, New York: Mouton de Gruyter.

Sneddon, James N. (1996). *Indonesian: A Comprehensive Grammar.* New York: Routledge.

Soe, Myint (1999). A Grammar of Burmese. PhD dissertation, University of Oregon.

Sohn, Ho-Min (1994). *Korean.* London: Routledge.

Sohn, Ho-Min (2009). The semantics of clause linking in Korean. In: *The Semantics of Clause Linking: A Cross-Linguistic Typology.* Eds. R.M.W. Dixon and Alexandra Y. Aikhenvald. Oxford: Oxford University Press. 285–317.

Soukka, Maria (2000). *A Descriptive Grammar of Noon – A Cangin Language of Senegal.* München: Lincom Europa.

Spronck, Stef (2012). Minds divided: Speaker attitudes in quotatives. In: *Quotatives: Cross-Linguistic and Cross-Disciplinary Perspectives.* Eds. Isabelle Buchstaller and Ingrid van Alphen. Amsterdam, Philadelphia: John Benjamins. 71–116.

Sridhar, Shikaripur N. (1976). Dative subjects. In: *Papers from the Twelfth Regional Meeting, Chicago Linguistics Society.* Eds. Salikoko S. Mufwene, Carol A. Walker and Sanford B. Steever. Chicago: Chicago Linguistics Society. 582–593.

Stassen, Leon (1985). *Comparison and Universal Grammar.* Oxford: Blackwell.

Stefanowitsch, Anatol and Stefan Th. Gries (2003). Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics 8.*2: 209–243.

Stiebels, Barbara (2007). Towards a typology of complement control. In: *Studies in Complement Control (ZAS Papers in Linguistics 47).* Ed. Barbara Stiebels. ZAS Berlin. 1–80.

Sundaresan, Sandhya and Thomas McFadden (2009). A selectional account of DP distribution. Paper presented at *Generative Grammatik des Südens*, Leipzig, May 23, 2009.

Suttles, Wayne (2004). *Musqueam Reference Grammar*. Vancouver, Toronto: University of British Columbia Press.

Szabolcsi, Anna (2009). Overt nominative subjects in infinitival complements in Hungarian. In: *Approaches to Hungarian 11*. Eds. Marcel den Dikken and Robert M. Vago. Amsterdam, Philadelphia: John Benjamins. 251–276.

Szent-Iványi, Béla (1995). *Der ungarische Sprachbau*. Hamburg: Buske.

Szmrecsanyi, Benedikt (2005). Language users as creatures of habit: A corpus-based analysis of persistence in spoken English. *Corpus Linguistics and Linguistic Theory* 1.1: 113–150.

Szmrecsanyi, Benedikt and Bernd Kortmann (2009). The morphosyntax of varieties of English worldwide: A quantitative perspective. *Lingua* 119.11: 1643–1663.

Tabor, Whitney and Michael K. Tanenhaus (1999). Dynamical models of sentence processing. *Cognitive Science* 23.4: 491–515.

Talmy, Leonard (2000). *Toward a Cognitive Semantics. Vol. I: Concept Structuring Systems.* Cambridge, MA: MIT Press.

Tamura, Suzuko (2000). *The Ainu Language.* Tokyo: Sanseido.

Taweesak, Kunyot (1984). General Characteristics of Hmong Njua Grammar. M.A. thesis, Mahidol University.

Taylor, Charles (1977). Presuppositional clauses in Nkore-Kiga. In: *Language and Linguistic Problems in Africa: Proceedings of the VII Conference on African Linguistics.* Eds. Paul F.A. Kotey and Haig Der-Houssikian. Columbia: Hornbeam Press. 249–255.

Taylor, Charles V. (1985). *Nkore-Kiga.* London: Croom Helm.

Terrill, Angela (2003). *A Grammar of Lavukaleve.* Berlin, New York: Mouton de Gruyter.

Thompson, James J. (2012). Syntactic Nominalization in Halkomelem Salish. PhD dissertation, University of British Columbia.

Thompson, Laurence C. (1965). *A Vietnamese Grammar.* Seattle: University of Washington.

Thompson, Sandra A. (2002). 'Object complements' and conversation: Towards a realistic account. *Studies in Language* 26.1: 125–164.

Thompson, Sandra A. and Elizabeth Couper-Kuhlen (2005). The clause as a locus of grammar and interaction. *Discourse Studies* 7.4–5: 481–505.

Thompson, Sandra A. and Anthony Mulac (1991a). A quantitative perspective on the grammaticization of epistemic parentheticals in English. In: *Approaches to Grammaticalization. Vol. II.* Eds. Elizabeth Traugott and Bernd Heine. Amsterdam, Philadelphia: John Benjamins. 313–329.

Thompson, Sandra A. and Anthony Mulac (1991b). The discourse conditions for the use of complementizer *that* in conversational English. *Journal of Pragmatics* 15: 237–251.

Thompson, Sandra A., Joseph S. Park & Charles N. Li (2006). *A Reference Grammar of Wappo.* Berkeley etc.: University of California Press.

Tikkanen, Bertil (2001). Converbs. In: *Language Typology and Language Universals: An International Handbook.* Eds. Martin Haspelmath, Ekkehard König, Wulf Oesterreicher and Wolfgang Raible. Berlin, New York: de Gruyter. 1112–1123.

Timberlake, Alan (1977). Reanalysis and actualization in syntactic change. In: *Mechanisms of Syntactic Change.* Ed. Charles N. Li. Austin, London: The University of Texas Press. 141–177.

Tomasello, Michael (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition.* Harvard: Harvard University Press.

Tomlin, Russell (1986). *Basic Word Order: Functional Principles.* London: Croom Helm.

Trask, Robert L. (1997). *The History of Basque.* London, New York: Routledge.

Traugott, Elizabeth (1972). *A History of English Syntax.* New York: Holt, Rinehart & Winston.

Tsunoda, Tasaku (1981). *The Djaru Language of Kimberley, Western Australia.* Pacific Linguistics, Series B 78. Canberra: The Australian National University.

Underhill, Robert (1976). *Turkish Grammar.* Cambridge, MA: MIT Press.

Van der Voort, Hein (2004). *Kwazá.* Berlin, New York: Mouton de Gruyter.

Van Geert, Paul (2008). The dynamic systems approach in the study of L1 and L2 acquisition: An Introduction. *The Modern Language Journal*, 92.2: 179–199.

Van Gijn, Rik (2006). A Grammar of Yurakaré. PhD dissertation, Radboud Universiteit Nijmegen.

Van Gijn, Rik (2011). Semantic and grammatical integration in Yurakaré subordination. In: *Subordination in Native South American Languages.* Eds. Rik van Gijn, Katharina Haude and Pieter Muysken. Amsterdam, Philadelphia: John Benjamins. 169–192.

Van Klinken, Catharina Lumien (1999). *A Grammar of the Fehan Dialect of Tetun – An Austronesian Language of West Timor.* Canberra: Australian National University.

Van Lier, Eva (2009). *Parts of Speech and Dependent Clauses: A Typological Study.* Utrecht: LOT Publications.

Van linden, An and Jean-Christophe Verstraete (2011). Revisiting deontic modality and related categories: A conceptual map based on the study of English modal adjectives. *Journal of Pragmatics* 43: 150–163.

Van Valin, Robert D. Jr. (1977). Aspects of Lakota Syntax: A Study of Lakota (Teton Dakota) Syntax and its Implications for Universal Grammar. PhD dissertation, University of California at Berkeley.

Van Valin, Robert D. Jr. (1984). A typology of syntactic relations in clause linkage. *Berkeley Linguistics Society* 10: 542–558.

Van Valin, Robert D. Jr. and Randy LaPolla (1997). *Syntax: Structure, Meaning, Function.* Cambridge: Cambridge University Press.

Verhoeven, Elisabeth (2010). Agentivity and stativity in experiencer verbs: Implications for a typology of verb classes. *Linguistic Typology* 14.2: 213–251.

Verstraete, Jean-Christophe (2008). The status of purpose, reason and intended endpoint in the typology of complex sentences. *Linguistics* 46: 757–788.

Vogt, Hans (1971). *Grammaire de la Langue Géorgienne.* Oslo: The Institute for Comparative Research in Human Culture.

von Eye, Alexander (1990). *Introduction to Configural Frequency Analysis.* Cambridge: Cambridge University Press.

Wakasa, Motomichi (2008). A Descriptive Study of the Modern Wolaytta Language. PhD dissertation, University of Tokyo.

Wang, William (1969). Competing changes as a cause of residue. *Language* 45: 9–25.

Wang, Yu-Fang (2006). The information structure of adverbial clauses in Chinese discourse. *Taiwan Journal of Linguistics* 4: 49–88.

Warner, Anthony (1982). *Complementation in Middle English and the Methodology of Historical Syntax.* London, Canberra: Croom Helm.

Wash, Suzanne (1999). On the structure and function of relative clauses in Barbareño Chumash. Unpublished Ms., University of California at Santa Barbara.

Wash, Suzanne (2001). Adverbial Clauses in Barbareño Chumash Narrative Discourse. PhD dissertation, University of California at Santa Barbara.

Wasow, Thomas A. (2002). *Postverbal Behavior.* Stanford, CA: CSLI Publications.

Watkins, Laurel J. (1984). *A Grammar of Kiowa.* Lincoln: University of Nebraska Press.

Watters, James K. (1985). Studies of the Tepehua Verb: A Functional Approach. M.A. thesis, University of California at Berkeley.

Watters, James K. (1988). Topics in Tepehua Grammar. PhD dissertation, University of California at Berkeley.

Weber, David J. (1989). *A Grammar of Huallaga (Huánuco) Quechua.* Berkeley: University of California Press.

Weber, David J. (1994). *Relativizacion y Clausulas Nominalizadas en el Quechua Huallagiono.* Yarinacocha: Instituto Lingüístico de Verano.

Weinert, Regina (2012). Complement clauses in spoken German and English: Syntax, deixis and discourse-pragmatics. *Folia Linguistica* 46.1: 233–265.

Whitehead, Carl R. (2004). A Reference Grammar of Menya. PhD dissertation, University of Manitoba.

Wiemer, Björn and Veronika Kampf (2013). Gesten und Mimik als semiotische Substitute für Sprechakte oder umgekehrt? Zur speziellen Verwendung einiger Reformulierungsmarker im Bulgarischen. *Съпоставително езикознание XXXVIII* 2/3. 153–191.

Wildgen, Wolfgang (1984). *Sprache als selbstorganisierendes System.* Trier: L.A.U.T.

Windfuhr, Gernot L. (1979). *Persian Grammar: History and State of Its Study.* The Hague: Mouton.

Wittgenstein, Ludwig (1953). *Philosophical Investigations.* New York: Macmillan.

Witzlack-Makarevich, Alena (2010). Typological Variation in Grammatical Relations. PhD dissertation, University of Leipzig.

Woollams, Geoff (1996). *A Grammar of Karo Batak.* Canberra: Pacific Linguistics.

Wrona, Janick (2005). The Modern Japanese complementisers *no* and *koto* and their Old Japanese precursors: A diachronic explanation for free variation. In: *Oxford University Working Papers in Linguistics, Philology & Phonetics.* Ed. Anna McNay. Oxford: The Centre for Linguistics and Philology. 119–136.

Wrona, Janick (2007). The early history of *no* as a nominalizer. Paper presented at the 7th NAJAKS conference, Copenhagen, August 2007. Available online at <http://ling.bun.kyoto-u.ac.jp/~wrona/History%20of%20no%20(NAJAKS).pdf> (as at October 2011).

Wrona, Janick (2008). *The Old Japanese Complement System.* Leiden: Brill.

Wu, Tong (2011). The syntax of prenominal relative clauses: A typological study. *Linguistic Typology* 15.3: 569–623.

Yamashita, Hiroko and Franklin Chang (2001). Long before short preference in the production of a head-final language. *Cognition* 81: B45–B55.

Yang, Byong-Seon (1994). Morphosyntactic Phenomena of Korean in Role and Reference Grammar: Psych-Verb Constructions, Inflectional Verb Morphemes, Complex Sentences and Relative Clauses. PhD dissertation, State University of New York at Buffalo.

Yap, Foong Ha and Stephen Matthews. (2008). The development of nominalizers in East Asian and Tibeto-Burman languages. In: *Rethinking Grammaticalization: New Perspectives.* Eds. María José López-Couso and Elena Seoane, in collaboration with Teresa Fanego. Amsterdam, Philadelphia: John Benjamins. 309–341.

Yap, Foong Ha, Karen Grunow-Hårsta and Janick Wrona (2011). Introduction: Nominalization strategies in Asian languages. In: *Nominalization in Asian Languages: Typological and Diachronic Perspectives.* Amsterdam, Philadelphia: John Benjamins. 1–57.

Ylikoski, Yussi (2003). Defining non-finites: Action nominals, converbs and infinitives. *SKY Journal of Linguistics* 16: 185–237.

Yue, Anne O. (2003). Chinese dialects: Grammar. In: *The Sino-Tibetan Languages.* Eds. Graham Thurgood and Randy J. LaPolla. London: Routledge. 84–125.

Zifonun, Gisela, Ludger Hoffman, Bruno Strecker and Joachim Ballweg (1997). *Grammatik der deutschen Sprache. Band 3.* Berlin, New York: de Gruyter.

Zipf, George (1935). *The Psychobiology of Language: An Introduction to Dynamic Philology.* Cambridge, Mass.: MIT Press.

Zúñiga, Fernando (2006). *Deixis and Alignment. Inverse Systems in Indigenous Languages of the Americas.* Amsterdam, Philadelphia: John Benjamins.

Ehrenwörtliche Erklärung

Ich versichere hiermit, dass ich die vorliegende Dissertation selbst angefertigt habe. Alle von mir verwendeten Hilfsmittel und Quellen sowie unterstützende Personen, die dankbarerweise als sprachliche Informanten oder Bereitsteller analytischer Hilfsmittel fungiert haben, sind in der Arbeit angegeben. Alle Textabschnitte fremder Autoren sind als solche kenntlich gemacht. Die Hilfe eines Promotionsberaters wurde nicht in Anspruch genommen; Dritte haben weder unmittelbar noch mittelbar geldwerte Leistungen von mir für Arbeiten, die im Zusammenhang mit dem Inhalt der vorgelegten Dissertation stehen, erhalten. Ich versichere weiterhin, dass die Dissertation noch nicht als Prüfungsarbeit für eine wissenschaftliche Prüfung eingereicht wurde. Weder wurde eine gleiche, eine in wesentlichen Teilen ähnliche oder eine andere Abhandlung bei einer anderen Hochschule als Dissertation eingereicht. Die geltende Promotionsordnung der Friedrich-Schiller-Universität Jena ist mir bekannt.

Weimar, 08.05.2014