PATTERNS OF MORPHOSYNTACTIC AND FUNCTIONAL DIVERSIFICATION IN THE USAGE OF COGNATE VERBS IN INDO-IRANIAN

by

SHAHAR BARUCH SHIRTZ

A DISSERTATION

Presented to the Department of Linguistics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2017

DISSERTATION APPROVAL PAGE

Student: Shahar Baruch Shirtz

Title: Patterns of Morphosyntactic and Functional Diversification in the Usage of Cognate Verbs in Indo-Iranian.

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Linguistics by:

Cynthia Vakareliyska	Chairperson
Scott DeLancey	Core Member
Spike Gildea	Core Member
Doris Payne	Core Member
Zhuo Jing-Schmidt	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2017.

DISSERTATION ABSTRACT

Shahar Baruch Shirtz

Doctor of Philosophy

Department of Linguistics

June 2017

Title: Patterns of Morphosyntactic and Functional Diversification in the Usage of Cognate Verbs in Indo-Iranian.

This is a study of processes of structural and functional diversification of the uses of three cognate verbs across the Indo-Iranian language family: “do/make”, “be/become”, and “give”. First, this study identifies over sixty distinct construction types in which these verbs are used, including complex predicate constructions, nominal predication constructions, serial verb constructions, and several distinct auxiliary constructions. Since the sets of verbs studied here are cognates, and share a common source, crosslinguistic differences in their uses are the result of grammatical change, and especially shared and parallel innovations of similar uses.

Then, this study presents a taxonomy of different complex predication types with “do/make”, and shows that there are general patterns in the deployment of different types of complex predication to express different types of situations. These patterns exhibit “transitivity prominence” previously identified by typologists with “heavy” or “lexical” verbs. This study then shows that these patterns are the result of several distinct pathways of grammatical change, often motivated by analogy to existing constructions, giving raise to different types of N-V complex predication constructions.

Then, this study shows that despite the fact that Indo-Iranian speakers can potentially deploy distinct constructions to encode each of the six nominal predication functions, sets of such functions are often co-expressed by the same structural coding means, especially clauses with cognate "be/become" verbs. This study uses a novel method, based on bipartite network graphs, to compare of the degree to which nominal predication functions are co-expressed in different languages.

Finally, this study shows that the three sets of cognate verbs are more likely to be used similarity within branches and subbranches of Indo-Iranian than across branches. The scope of this branches, however, is different for different verbs: "do/make" and "give" behave more similarly in languages which belong to the same major branch, Iranian or Indo-Aryan, but "be/become" clusters are at different levels of subbranching. This is the result of the different types of innovations attested with these verbs: reanalysis and actualization motivated by analogy with "do/make" and "give", and metaphorical and metonymy extensions with "be/become".

CURRICULUM VITAE

NAME OF AUTHOR: Shahar Baruch Shirtz

GRADUATE AND UNDERGRADUATE SCHOOL ATTENDED:

University of Oregon, Eugene

Hebrew University of Jerusalem, Jerusalem

DEGREES AWARDED:

Doctor of Philosophy, Linguistics, 2017, University of Oregon, Eugene

Master of Arts, Linguistics, 2010, Hebrew University of Jerusalem, Israel

Bachelor of Arts, Linguistics and Cognitive Science, 2006, Hebrew University of Jerusalem, Israel

AREAS OF SPECIAL INTEREST:

Linguistic Typology (especially discourse functions, complex predication, nominal predication)

Historical and Comparative linguistics (especially the diversification of cognate morpho-syntax)

Corpus / Usage Based Typology

Construction(al models of) Grammar

Event structure

Discourse grammar

PROFESSIONAL EXPERIENCE:

Graduate Teaching Fellow, University of Oregon, September 2010 – June 2017

PUBLICATIONS:

Shirtz, Shahar. 2016. "Indirect Participants as core arguments in Middle Persian". In: Jila Ghomeishi, Carina Jahani, and Agnès Lenepvue-Hotz (eds), *Further topics in Iranian linguistics*. Cahiers de Studia Iranica 51. 175-194. Leuven: Peeters.

Payne, Doris L & **Shirtz Shahar** (eds). 2015. *Beyond Aspect: The Expression of Discourse Functions in African Languages.* Typological Studies in Language 109. Amsterdam/Philadelphia: John Benjamins. vii, 321 pp.

Shirtz, Shahar & Payne, Doris L. 2015. "Discourse structuring and typology: How strong is the link with aspect?". In: Doris L Payne and **Shahar Shirtz** (eds), *Beyond Aspect: The Expression of Discourse Functions in African Languages*. Typological Studies in Language 109. 1-22. Amsterdam/Philadelphia: John Benjamins.

Shirtz, Shahar & Payne, Doris L. "Nominal heads in Maa". In: Ọlanikẹ Ọla Orie and Karen W. Sanders (eds), *Selected Proceedings of the 43rd Annual Conference on African Linguistics: Linguistic Interfaces in African Languages*. 207-221. Somerville MA: Cascadilla Press. (Available at: http://www.lingref.com/cpp/acal/43/paper2969.pdf)

ACKNOWLEDGMENTS

It is easier to find the words for a sound grammatical description than it is to find the words to express what I wish to express here, so I will not pretend like this does it.

First and foremost, I wish to thank my advisor, Cynthia Vakareliyska and my committee members, Doris Payne, Spike Gildea, Scott DeLancey, and Zhuo Jing-Schmidt. Without their criticism and encouragement this would not have been possible. I also wish to thank Tyler Kendell, Volya Kapatsinski, and Eric Pederson for their help, advice, and critique along the way, and to Shaul Shaked, who hired my as a research assistant in Jerusalem years ago and (inadvertently?) introduced me to the world of Indo-Iranian languages.

In Oregon, I presented some of the ideas expressed here in the Linguistics department colloquium, and Fieldling / LingDocRev and CogGroup meetings. These meetings proved extremely helpful in shaping and reshaping my ideas. Just as important, mentally and professionally, were discussions over beverages with fellow colleagues: Linda Konnerth, Chris Doty, Jaime Peña, Danielle Barth, Brian Butler, Prakaiwan Vajrabhaja, Ted Bell, Zoe Tribur, Paul Olejarczuk, Matt Stave, Manuel Otero, Zalmai "Zeke" Zahir, Krishna Boro, Julia Trippe, Amos Teo, Sara Pacchiaroti, Becky Paterson, Charlie Farrington, Jason McLarty, Marie-Caroline Pons and many others. Without them, this entire experience would have been worse.

I wish to thank my family in Israel, my parents, grandparents, and brothers (now with a sister-in-law and a nephew!), for being supportive even if they didn't really understand why I do the things that I do. Finally, I also wish to thank Jessica (and

Vynnie, Rock, Riley, and Remy), my family in Eugene, Oregon, without whom it would not have been worth it.

To my family, here and there.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

This dissertation examines the functional and structural diversity of usages attested with three sets of cognate verbs, "do/make", "be/become", and "give", across the Indo-Iranian language family. Crosslinguistically, these verbs are the lexical sources of several grammatical functions, and across the Indo-Iranian language family they are used in a variety of functions, including as the main verbal predicate in clauses expressing many different types of states and events, and as auxiliary verbs in constructions expressing several distinct grammatical categories.

Since the verbs analyzed here are cognate verbs, crosslinguistic differences in their usage point to differences in the historical processes which were active at different times across the family. The diversity of functions and usages of these three verbs raise many questions about the analysis of synchronic variation and the processes which led to it. This study asks, and answers, several sets of these questions.

The first set of questions is taxonomic in nature: what are the attested usages of these three sets of cognate verbs across Indo-Iranian? While the main focus of this question is descriptive, it requires that the description of language-specific usages of "do/make", "be/become", and "give" be drawn up using comparative terms alone, with no reference to language-specific constructions. This is done by defining a set of construction types as bundles of comparative concepts (following Hockett 1955[1], Stassen

[1] In a paper dealing with attribution and apposition, Hockett 1955 uses the term "construction type" in a very similar way to the definitions of "strategy" comparative concepts (see below). The use of this term in this study is slightly different from Hockett's use of the term.

1985:14, Haspelmath 2010, Croft 2016 *inter alia*), and using them to compare the uses of these verbs in crosslinguistically.

The next set of questions deals with "do/make" in Noun-Verb (henceforth N-V) complex predicate constructions, one of the most frequent uses of "do/make" in the data. N-V complex predicates are constructions in which the main verbal predicate is accompanied by a nominal element which encodes the situation the clause expresses; and the structure of the clause, especially its argument structure, is determined jointly by that noun and the verb. I ask, following Mohanan 1994, 1997 and Haig 2002, what the attested types of N-V complex predication are in the data, and whether there is an aggregate overall relationship between the type of situation a clause expresses and the argument structure of that clause. Then, I ask what the diachronic processes are which lead to the observed variation in argument structure constructions with N-V complex predicates in the data.

The next questions concentrate on the most common use of "be/become" in the data, as the main verbal predicate in clauses encoding the six nominal predication functional domains: equation, predicate attribute, proper inclusion, predicative possession, the predicate locative, and the existential. I ask what are the patterns where these six functional domains are co-expressed by clauses using the same configuration of structural coding means. I then propose a novel method, based on bipartite networks for enabling a crosslinguistic comparison of co-expression patterns.

The final set of questions asks whether the aggregate usage patterns of "do/make", "be/become" and "give" are more similar within subfamilies of Indo-Iranian than across subfamilies. These questions are answered by utilizing Neighbor-Net, a

quantitative distance based method, to compare the aggregate usages of these three verbs across Indo-Iranian.

The main data source for this study is described in detail in Chapter II. For now, suffice it to note that the data is composed of a set of published primary naturalistic texts in a sample of sixteen Indo-Iranian languages, divided between the Indo-Aryan and Iranian branches of Indo-Iranian. These languages include one Old Indo-Aryan language (Sanskrit), several Middle Indo-Iranian languages (Bactrian, Sogdian, and Middle Persian from the Iranian branch; Pali from the Indo-Aryan branch), and several Modern Indo-Iranian languages (Early Judeo-Persian, Sivandi, Ziyarat Mazandarani, and Gorani from the Iranian branch; Hindi, Gujarati, Kotia Oriya, Kupia, Nagamese, Palula, and Darai from the Indo-Aryan branch).

All occurrences of "do/make", "be/become", and "give" attested in these texts are extracted and analyzed. The patterns attested there are further analyzed by quantitative tools such as Neighbor-Net and bipartite networks. While Neighbor-Nets have been gaining popularity in comparative linguistics over the last decade, the use of bipartite networks as a tool in comparative and descriptive linguistics is more novel. This study then, approaches the questions asked by using primary texts as data and analyzing the result using established and novel quantitative methods. As a result, the answers this study gives to the questions it asks less susceptible to inadvertent cherry picking.

Consider, for example, that in many Indo-Iranian languages predicative possession is expressed by the same means as the predicate locative, and such instances have been used to argue for a privileged relationship between possessors and locations. In many of these languages, however, there are further constructions expressing predicative

possession which are distinct from constructions expressing the predicate locative. Thus, the overlap between the means expressing predicative possession and those expressing the predicate locative is often partial and crosslinguistically variable. It is difficult to measure the degree of this overlap without using primary textual data, and without using the bipartite networks as a model for the data. These two, then, allow for developing a more realistic picture of the expression of these functions, and produces results which are more readily testable and reproducible.

The remainder of this introduction sketches the orientation of this study. First, Section 1.1 explains the differences between descriptive categories and comparative concepts, and the implications of the view that linguistic categories are language-specific categories for the questions posed in this study. Section 1.2 summarizes the main mechanisms of morphosyntactic change that are assumed in this study, and their interaction with a constructional model of language. Section 1.3 sketches the crosslinguistic variation in coding properties attested across Indo-Iranian, and the approach of this study to argument structure constructions. The introduction concludes with a brief overview of the remaining chapters in this study.

1.1 Descriptive Categories, Comparative Concepts

The approach of this study is based on the general model of Construction Grammar as articulated in Goldberg 1995, 2006, and, more specifically, on Croft's Radical Construction Grammar model (Croft 2001). The analyses in this study assume that the basic unit of grammar is that of a construction: a form-meaning, or form-function, correspondence, composed of a semantic, or functional, pole; a syntactic, or formal, pole;

and a symbolic link or links between them (Croft 2001:18; the term "pole" is due to Steels 2011). Thus, a lexical item such as *barf* 'snow' in New Persian is a construction; a morpheme such as the Nagamese Plural[2] marker *-khan* is a construction; and an abstract collocation composed of several functionally and syntactically related constituents, such as the English Double Object construction [DP V DP DP] (cf. *she gave me the book*), is also a construction.

The analyses provided in this study do not assume a specific formal model of construction grammar such as Sign Based Construction Grammar (SBCG, e.g., Boas and Sag 2012) or Fluid Construction Grammar (FCG, e.g., Steels 2011). I believe that all the analyses provided here could potentially be formalized within one of a number of construction-based formal models, as well as other closely-related constraint-based models such as Head-Driven Phrase Structure Grammar (HPSG, e.g., Müller 2016:255-302).

Constructions are seen here, first and foremost, as language-specific grammatical entities composed of form-plus-function which are necessary for the description of specific languages. The logic behind this view of constructions has been articulated in many venues (e.g., Croft 2001, 2009, 2014; see also papers in Boas 2010). Here I wish to briefly demonstrate it with two examples: the major bivalent constructions with "do/make" in Hindi and Kupia, and the plural markers in Middle Persian and Nagamese.

Abstract clause-level constructions, including argument structure constructions, cannot be directly equated across languages. Consider, for example, the two clauses in

---

[2] In this study, following the usual convention in typological studies, the first letter of words referring to language-specific grammatical categories, such as the Nagamese Plural, is capitalized. The non-capitalized version, i.e., "plural," refers to the general grammatical category (comparative concept) of the plural.

(1.1a-b) below, from Hindi and Kupia, respectively. In both clauses, the main verbs are cognate "do/make" verbs, accompanied by two arguments. The first argument expresses the microrole of doer, and the second is the interrogative word "what". At face value, it seems the two constructions are directly comparable: a bivalent usage of "do/make".[3]

The internal structure of each of these constructions, however, is quite distinct. In example (1.1a), the doer argument is flagged by the Ergative marker *=ne* and the verb indexes the interrogative "what". In example (1.1b), neither argument is flagged by a marker, and the verb indexes the doer argument in person, gender, and number. Thus, the internal relationship between the different constituents is quite different in the two constructions.

(1.1a) *tum=ne kya kiyā?*
2SG=ERG what do.PRF.FSG
'What did you do?' (Hindi, Premchand 2017[1936]:3.22)

(1.1b) *guru kicco ker-l-o?*
teacher what do-PST-MSG
'What did the teacher do?' (Kupia, Christmas & Christmas 1973a:75)

One could argue that instead of the Hindi Ergative construction in (1.1a), I could have used the Hindi Accusative construction, where the internal relationships between constituents are quite similar to (1.1b). This is true, but it highlights the language-specific nature of constructions: that is, the distribution of the Hindi Accusative construction with

[3] Below, and especially in Chapter IV of this dissertation, I argue that these two constructions are indeed comparable, but only if we define them in terms of comparative concepts.

"do/make" is constrained in terms of Tense, Aspect, Mode (henceforth TAM), whereas the distribution of the Kupia Accusative construction in (1.1b) is not. Thus, while the internal structures of the Hindi Accusative construction and the Kupia Accusative construction are similar, their distributional properties are different, and hence they too cannot be argued to be the *same* construction.

The same problem is encountered when considering morpheme-sized constructions which express "the same" grammatical category. Consider, for example, the plural markers in Middle Persian and Nagamese, which differ in their distribution across different lexical classes. In Middle Persian, the Plural marker *-ān* is found attached to lexical nouns, as in *mardān* 'men', and, less frequently, to adjectives, as in *gursagān* 'hungry (people)'. It is never found attached to independent pronouns such as first and second person pronouns; for example, the Middle Persian first person plural form is *amāh* 'we'. The Nagamese Plural marker *-khan,* on the other hand, is found attached both to lexical nouns, such as *duthkhan* 'angels', and to personal pronouns, such as *apunikhan* '2PL' or *amikhan* '1PL', but, as far as the data analyzed here shows, not to adjectives. Thus, the distribution of the plural markers differs between Middle Persian and Nagamese.

Further, across the documented history of Middle Persian, the grammar associated with the suffix *-ān* changed quite dramatically. In the data used for this dissertation, Late Zoroastrian Middle Persian[4] *-ān* functions as the general plural marker alongside the suffix *-īha* , which expresses "individual plurality"[5]. In earlier stages of Middle Persian

[4] Roughly texts written between the 8th and 11th centuries; see Josephson 2016 for a brief overview.
[5] Skjærvø (2009:205) demonstrates this difference by *kōfān* 'mountains' vs. *kōfīhā* 'the various / individual mountains'. I remain agnostic about this description.

(cf. inscriptional Middle Persian), *-ān* conflated case and number. In texts from this period, *-ān* expresses both case and number: *mard* 'man' was the form used for the Direct case[6] Singular and Plural, and the Oblique case singular, and *mardān* 'men' for the Oblique Plural only.[7] So the Plural marker *-ān* is a language specific construction even within different types (or stages) of Middle Persian.

The examples above, together with many other examples in the linguistic literature (see, for example, Croft 2001, Haspelmath 2010 for overview and examples) illustrate the language-specific nature of constructions. This view of construction as language-specific grammatical entities is a specific instance of the more general phenomenon that linguistic categories are language-specific, a position which has been argued in Linguistics since at least Boas (1911), if not before. This nature of linguistic categories poses an obvious challenge to linguists who are interested in comparative questions, historical or typological.

Tackling this issue, many linguists, as early as Greenberg (1963; also Hockett 1955), have conducted crosslinguistic comparisons using terms defined without reference to any language-specific categories and without necessarily assuming the crosslinguistic validity of constructions and categories. Haspelmath (2010) offers a clear characterization of this strategy for crosslinguistic comparison[8], distinguishing descriptive categories from comparative concepts. In Haspelmath's terminology,

---

[6] Case terminology in Indo-Iranian, and in this study, is explained in section 1.3.3 below. For now, the Oblique case in Early Middle Persian is treated as the case form used to flag lexical nouns which function as P arguments when the verb is in the present tense, A arguments when the verb is in the past tense, and lexical nouns governed by a preposition. See below for definitions of A arguments.

[7] With a subset of lexical nouns expressing kin relations, *-ān* interacts with another suffix, *–ar,* resulting in an agglutinative structure.

[8] He articulates this position against the universalist position in Newmeyer (2007) that all languages can use grammatical categories from a universal inventory.

DESCRIPTIVE CATEGORIES are language-specific categories which are necessary for the description of a given language. These can be language-specific constructions, like the Nagamese Plural marker or the Hindi Ergative construction, or more abstract linguistic categories such as parts of speech (for discussion, see, for example, Croft 2001:63-107).

In order to provide a solid basis for crosslinguistic comparison, Haspelmath (2010) argues for the method employed by many typologists (albeit implicitly by some) of defining COMPARATIVE CONCEPTS, i.e., concepts devised for crosslinguistic comparison by linguists asking comparative questions. According to Haspelmath's definition, comparative concepts are not a part of the grammatical system of specific languages, and hence they do not need to be a part of the linguistic knowledge of speakers and are not necessarily represented in the minds of speakers of a particular language. Haspelmath (2010) and Croft (2016) identify several distinct types of comparative concepts. Some are purely semantic, or functional, and are defined on the basis of their function. As an example of a set of purely semantic comparative concepts, Croft 2016 proposes stimuli sets or questionnaires designed for crosslinguistic study of semantic domains, such as the stimuli set designed by Bowerman and Pederson (1992) for the study of spatial relations. Each stimulus in their set is composed of a figure and a ground (e.g., a cup on a table), and can be thought of as a purely semantic comparative concept. Other functional comparative concepts are categories such as the passive, as defined by Givón (2001), or each of the six nominal predication functional domains, as defined in Stassen 1997:11-21 and Payne 1997:111-128.

Most comparative concepts constructed by linguists refer to both function and form, and are usually called HYBRID COMPARATIVE CONCEPTS (e.g., Stassen 1985:14,

Haspelmath 2010, Croft 2016). This is the case for most of the comparative concepts presented in this study: they refer to the function of a specific construction and its components, and to properties of their grammatical form. This type of hybrid comparative concept is called a "strategy" by Croft, who defines it as follows:

> "a construction in a language (or any language), used to express a particular combination of semantic structure and information packaging function, that is further distinguished by certain characteristics of grammatical form that can be defined in a crosslinguistically consistent fashion" (2014:537).[9]

Consider, for example, the clauses in (1.1a-b) above, where "do/make" is accompanied by two arguments. In Chapter IV, I argue that these clauses fit common construction type, attested in many of the languages analyzed here. I define this construction type as follows:

> a bivalent clause construction where the main verbal predicate is a cognate of PIE **k^{w}er* "do/make", and in which one argument, coded as an A argument, expresses the doer microrole and another argument, coded as a P argument, expresses an object the doer creates or a situation the doer is involved in.

[9] A similar approach has long been the common method of functional typology (e.g., Givón 1979, 2001): that is, first define a functional domain such as passive or causative, and then identify the different structural means employed to express this function crosslinguistically.

The terms “bivalent”, “A argument” and “P argument” are comparative concepts themselves (the definition of A and P arguments as comparative concepts is discussed in detail in Haspelmath 2011, and briefly below). The semantics of the A (“doer microrole”) and P (“an object the doer creates”) arguments, and the semantics of the clause, are functional comparative concepts. The inclusion of “cognates of PIE *$k^{w}er$”, a purely formal property, is motivated by the focus of this study: a comparison of the functional and structural diversification of sets of cognate verbs across Indo-Iranian. Thus, construction types like the ones defined in the quote immediately above are composed of a set of linked comparative concepts.

The definition of the construction type above makes no reference to language-specific constructions that code A and P arguments. Thus, it generalizes over different types of clause alignments such as the Hindi Ergative alignment and the Kupia Accusative alignment. The definition only references the fact that, for example, the doer microrole in this construction type is coded as an A argument. Whether A arguments have identical coding properties to S arguments, as in Kupia, or distinct ones, as in Hindi, makes no difference in the context of the definition above.

When “do/make”, “be/become”, or “give” functions as the main verbal predicate of the clause, definitions for specific construction types like the one given just above are modeled after definitions of argument coding frames (e.g., Haspelmath 2005, Malchukov et al. 2010), which spell out the coding properties and function of arguments that accompany the verb. When the verb functions as an auxiliary, the definition of construction types is based on the grammatical function encoded by the entire auxiliary construction.

The abbreviations S, A, P, T, and R have been used several times already in this introduction, but reference to their definitions has been made so far only made in passing. The definitions for these argument types as used here are based on the ones given by Comrie 1989 and Lazard 1994 (who uses X, Y, and Z, instead of S, A, and P), *inter alia*. In Comrie 1989, A and P arguments, for example, are defined as the most agent-like and most patient-like arguments, respectively, of a prototypical transitive clause with a verb such as "break." Lazard (1994, 2002, 2015) claims that crosslinguistically, the construction used with prototypical transitive verbs is often somehow privileged, and its use extends far beyond prototypical transitive events: e.g., that English *want* or *need* can occur with two arguments in the same argument frame as prototypical transitive verbs. The wide range of event types expressed by this construction led Lazard to refer to it as "the major biactantial construction", or the MBC (Lazard 2002:153). Haspelmath 2015 refers to a similar finding as "transitivity prominence."

Haspelmath 2011 offers detailed arguments showing that under such definitions, S, A, P, T, and R arguments are comparative concepts: they are identified in different languages based on the semantic terms "prototypical transitive clause", "most agent-like participant", and "most patient-like participant." It is this definition of A and P arguments as comparative concepts which allows us to use these terms in the definition of the construction type given above.

1.2 Types of Morphosyntactic Change in Construction Types.

The sets of cognate verbs analyzed in this dissertation, i.e., Indo-Iranian "do/make", "be/become", and "give", are used in a variety of functions defined here by "strategy"

comparative concepts and construction types. Across the data analyzed here, these three verbs function as main verbal predicates and as auxiliary verbs, and have sometimes grammaticalized further. Since these verbs are cognate, crosslinguistic differences in their usage are the result of grammatical change. Theoretically, these changes can be the result of innovations or a loss of uses that had already been present in some prior stage (e.g., Proto-Indo-Iranian). Most crosslinguistic differences identified in this study can be argued to be innovations, and apart from one or two cases, it seems that a complete loss of a construction type that was present in Proto-Indo-Iranian is infrequent.[10]

The mechanisms of morphosyntactic change that are assumed in this study, following Harris & Campbell (1995), are reanalysis, extension, and borrowing due to language contact. Briefly, Harris & Campbell (1995), following Langacker (1977), define REANALYSIS as a change in the underlying structure of a construction without changes in the overt expression of any constituent. Thus, when reanalysis takes place, the observed coding or behavioral properties of any constituent in the construction do not change. Changes in coding or behavioral properties are found in the next stage of the process, in which the reanalysis is actualized (using Timberlake 1977's term) by overt changes, such as changes in the flagging of participants from P arguments to oblique objects (see Chapter VII). EXTENSION, in Harris & Campbell 1995's terms, involves a change in the contexts in which a form can be deployed by speakers, without a corresponding change in the formal structure. For example, metaphorical uses of nouns such as *genius* or *asshole* to denote properties of referents instead of referring in themselves, as in *she is a genius* or

[10] Often a loss of a construction type is correlated with lexical replacement. For example, none of the instances of the Gujarati cognates of "give" in the data functions as a main verbal predicate. In those functions, it has been replaced by another verb, *ap-* 'give'.

I am an asshole can be thought of as an extension of their use without a subsequent change in phrase- or clause-level grammar (i.e., the two are still treated as nouns, not adjectives). Similar changes are found throughout Indo-Iranian and are analyzed in Chapters V and VIII.

Morphosyntatic change motivated by language contact or BORROWING is the result of social situations in which multilingual speakers use features originally associated with one language in social contexts in which they use the other language. Language contact has definitely been an active mechanism in the creation of new construction types across Indo-Iranian, and, in other instances, it instigated grammatical change across the family (see, e.g., Stilo 2004). Several Indo-Iranian languages analyzed here have been in sustained contact with each other (cf., the high social status of Persian across most of the Indo-Iranian speaking region) and with Turkic, Dravidian, or Tibeto-Burman languages. While some of the innovations identified here could be the result of languages contact, this study does not focus on contact-induced change, but primarily investigates language change which resulted from reanalysis and extension.

While these mechanisms of morphosyntactic change have not been specifically identified and formalized as a part of a constructional model of grammar, they can be incorporated into this kind of model. Such an incorporation is suggested by Barðdal and Gildea (2015; see also Barðdal and Eythórsson 2012), who identify three distinct loci of change in constructional models of language: the functional, or semantic, pole of the construction; its formal, or syntactic, pole; and the symbolic link between these two. Morphosyntactic change, including the rise of an innovative construction type, can involve changes in all three loci, or a combination of them.

For example, in all 16 languages analyzed in this study, cognate "be/become" verbs are used together with two unflagged NP constituents, in clauses expressing core nominal predication (equation, predicate attribute, and proper inclusion). In some languages, the semantics of the predicate NP sometimes extends to include place names (e.g., "France") and elements such as "home." This extension results in clauses expressing the predicate locative functional domain, and not core nominal predication. The construction type which expresses the predicate locative with "be/become" and two unflagged NPs is attested in several languages in the sample (e.g., Gorani, Kupia), and is analyzed in Chapters V and VIII. The emergence of this construction type is not accompanied by any changes in the coding properties of clausal constituents, and thus there are no changes in the syntactic, or formal, pole of the construction. The change has to do with the semantic pole alone: an extension of the semantics of the predicate NP and a respective change in the semantics of the whole construction.

The emergence of many other construction types involves changes in the coding means of some component, especially in flagging. The innovation of many N-V complex predicate constructions involves an innovation in the coding of the constituent expressing the affected participant, from being coded as a P argument to being coded as an oblique object. This is discussed in detail in Chapter VII. For example, the events described in (1.2a-b) are similar, and deal with assistance. In both, the verb is "do/make" and the type of situation is expressed by the NP *ayārīh* 'help, assistance', but in (1.2a), the affected participant is coded as a P argument by the unflagged NP *awēšān* 'they', and in (1.2b) as an oblique object by the PP *ō wehān* 'to good people'.

(1.2a) *ke awēšān ayārīh kun-ēd*
REL 3PL help do.PRS-3SG
'someone who helps people' (Middle Persian, PRDD 8J)

(1.2b) *ke pad zōr ud nērōg ī man ayārīh ō weh-ān nē kun-ē*
REL in strength and power LNK 1SG help to good-PL NEG do.PRS-2SG
'(you) who do not help good people by my strength and power' (Middle Persian, PRDD 26)

Even if two language-specific constructions belong to the same construction type, they are not necessarily cognate constructions. Instances of the same construction type can be, and often are, the result of parallel and distinct diachronic processes. Examples (1.3a-b), from Middle Persian and Kotia Oriya, are both instances of a construction type in which cognate "be/become" verbs are used in clauses expressing predicative possession in Heine's "Goal schema" ("X exists to/for Y"; Heine 1997). The two markers flagging the possessors in (1.3a-b) are not cognate, and while they share much of their functional range (e.g., both are used to flag recipients), they also differ quite a bit: for example, the Kotia Oriya *=ke* also flags P arguments, while the Middle Persian marker *rāy* does not. Thus the constructions illustrated in (1.3a-b) are not cognate. This construction type is analyzed in more detail in Chapter V.

(1.3a) *ud ōy wirāz rāy haft xwah būd h-ēnd*
and DEM wiraz to seven sister be.PST be.PRS-3PL
'and Wiraz had seven sisters' (Middle Persian, AWN 2.1)

(1.3b) *se dangɽ-i=ke pila oi ga-la*

DEM youngin-F=OBJ child be.NF go-PST.3SG

'The young girl had a child.' (Kotia Oriya, Gustafsson 1973a:194)

Some of the innovative uses of "do/make", "be/become", and "give" across the data are the result of GRAMMATICALIZATION, the gradual change in the functional or syntactic category of lexical items toward a more abstract one, often accompanied by a morpho-phonological reduction (e.g., Hopper and Traugott 1993, Bybee 2001, 2007). In much of the grammaticalization literature it has been shown that these changes take place in specific grammatical (i.e., syntactic and semantic) contexts (e.g., Hopper and Traugott 1993, Bybee 2007, Givón 2008, *inter alia*), and Haspelmath (2004:26) takes a decidedly constructional approach to grammaticalization, pointing to an increase of the internal dependency between constituents of a construction as a necessary property of grammaticalization.[11]

Many studies have treated grammaticalization processes as a cline, ranging from less abstract to more abstract categories (see, for example, studies in Trousdale and Traugott 2010, several of which, such as Hilpert 2010, assume a constructional model of grammar). This progression, from a fully lexical function to a more grammatical, or abstract, one has been often described as a GRAMMATICALIZATION CHAIN (Heine 1992:348-349, 2000). Crosslinguistically comparable lexical sources are often found at the more lexical end of such chains, and comparable functions at the other end. For

[11] Haspelmath 2004 uses this definition to argue that grammaticalization is unidirectional by definition, since changes which result in *weaker* ties would not fall under the purview of grammaticalization.

example, in their lexicon of grammaticalization, Heine and Kuteva (2002:117-120, 149-155) list verbs with "do/make" and "give" semantics as the source for at least five distinct functions apiece. "Be/become" verbs do not form a single entry, but the different copular functions of "be/become" verbs across Indo-Iranian are listed there as the source of almost two dozen functional domains.

Several studies, however, have called into question the primacy given to grammaticalization chains by some, arguing that grammaticalization processes occur one change at a time, and that these changes occur due to purely locally motivated processes of change involving the usual mechanisms of language change such as reanalysis and extension. Thus, grammaticalization chains are epiphenomena involving crosslinguistically common series of changes.

The view of grammaticalization chains as an epiphenomenon has been articulated, for example, in Givón 2013 where data from Hebrew, Spanish, and Tamil shows that the grammaticalization chain often offered for the development of the ethical dative is composed of locally motivated changes only. Similarly, Joseph (2014) argues, following Joseph & Pappas (2002), that the grammaticalization of the Greek Future marker *θa=* from *θɛlo* 'want' involves a rather complicated and messy route of reanalyses and extensions, and is not a simple and direct progression of semantic bleaching and phonological reduction.

Following these approaches, this study also views grammaticalization as a result of a series of reanalyses and extensions which occur vis-à-vis a particular construction, and not as a unique mechanism of change. When processes of change identified in this dissertation involve multiple steps (especially in Chapter VIII, but also occasionally in

Chapters IV through VI), each step is analyzed in terms of the traditional mechanisms of change.

1.3 Quick Survey of Argument Coding Means in Indo-Iranian.

This section briefly describes the variation in the main structural coding means across the Indo-Iranian language family that are considered in this study: relative word order, argument indexation, and noun flagging. While a thorough description of the crosslinguistic and intralinguistic variation in each of these coding means in Indo-Iranian could be the subject of a hefty monograph, the goal of this section is simply to orient readers not familiar with the structure of Indo-Iranian languages to the main coding means used by speakers of these languages for core arguments and oblique objects.

1.3.1 Word Order

Most descriptions of Indo-Iranian languages describe the basic word order of these languages as either verb-final or as SOV.[12] This is true not only for Indo-Iranian languages, but also for other major language families with which Indo-Iranian speakers have been in sustained contact, including Turkic, Dravidian, Tibeto-Burman, and Munda (for further discussion, see Masica 1976). Examples (1.4a-b), a ditransitive clause from Standard Oriya and a typical transitive clause from Gorani, illustrate the verb-final nature of Indo-Iranian:

[12] It is somewhat difficult to characterize the word order of Old Indo-Aryan languages, especially Sanskrit. Some linguists, e.g., Staal (1967), consider Sanskrit as having "free" word order, i.e., a language in which the word order is not correlated with grammatical function or semantic role.

(1.4a) *mũ tumɔ=ku bɔhi de-bi*
1SG 2SG=OBJ book give-FUT.1SG
'I shall give you the book.' (Standard Oriya, Masica 1991:333, glosses mine).

(1.4b) *tu min=it dī*
2SG 1SG=2SG see.PST
'You saw me.' (Gorani, Mahmoudveysi et al. 2012:39, glosses mine).

While verb-final word order is by far the most common one in texts, it is not entirely rigid. Across the family, the relative order of the arguments can change to express differences in their information status. This is shown in (1.5) from Hindi, where the A argument, flagged by the Ergative marker *=ne*, follows the R argument. Mohanan & Mohanan (1994) argue, supported by elicited examples, that the relative position of the verb is also variable.

(1.5) *mujhe to us=ne kutʃ.nahĩ diyā*
1SG.DAT TOP 3SG=ERG nothing give.PRF.FSG
'She gave me (absolutely) nothing.' (Hindi, Montaut 2012:191, my translation)

Apart from the non-verb-final word order that is correlated with changes in information status and information packaging, there are other systematic instances where the relative word order is not verb final. Across the Central Iranian languages, such as Sivandi and Gorani, goals of motion and recipients are often expressed by a post-verbal

phrase.[13] In (1.6a-b), from Sivandi, the verb is followed by an unflagged NP that expresses the goal of motion in (1.6a) and the recipient in (1.6b):

(1.6a) *Būzarjemehr kam kam ši kotowxāne*
Buzarjemehr little little go.PST school
'Slowly, Buzarjmeher went to school.' (Sivandi, Lecoq 1979:134)

(1.6b) *xūī=ra be-de-y ossād*
egg=OBJ IRR-give.PRS-1SG master
'I will give the egg to the master.' (Sivandi, Lecoq 1979:85)

The Sanskrit data for this study is in verse, a typical feature of Epic Sanskrit texts. In these, at least some instances of non-verb-final word order can be argued to be motivated by the metric properties of the specific meter of a foot. The word order in the Sanskrit data analyzed here, then, is not a reliable proxy of vernacular word order.

1.3.2 Argument Indexing[14]

While most verbal constructions across Indo-Iranian index at least one participant, indexation varies with respect to several variables: (a) verbs do not always index the same nominal categories; (b) indexation is not always achieved by the same type of coding means; and (c) indexation is not always a feature of a verbal clause construction.

[13] This may be an areal feature, since it is also found in Neo-Aramaic languages and some Turkish dialects; see Haig & Thiele (2014).
[14] In this study, I follow Croft's 2013 and Haspelmath's 2013 approach to argument indexation on the verb (see also Croft 2001:226-231).

Most verbal constructions across Indo-Iranian index a participant through suffixation. This is illustrated in (1.7a-b) from Pali and Middle Persian, where the verb indexes the A argument. In both examples, the verb indexes the A argument in person and number, but not in gender:

(1.7a) *atha kho bhagavā rājān-aṃ ... sand-essesi*
then but Lord.NOM king.ACC … instruct-AOR.3SG
'and then the Lord instructed … the king' (Pali, Mahāvagga II:1,4; Petesron 1998:90)

(1.7b) *ud drōd purs-ēnd ud āfrīn kun-ēnd ud ēn gōw-ēnd*
and greetings ask.PRS-3PL and praise do.PRS-3PL and DEM say.PRS-3PL
'They offered greetings, and gave praise, and said…' (Middle Persian, AWN 10.3)

In some constructions, especially in the Old and Middle Indo-Aryan languages such as Sanskrit and Pali, the verb indexes an argument in gender and number. This is illustrated in (1.8), an example of the Pali Gerundive construction, where the verb indexes the NP referring to the type of ceremony the priests are to perform.

(1.8) *tehi ... bhikkhūhi ... uposatho ... katabbo*
DEM.INST.MPL … monk.INST.PL … *uposatha*.NOM … do.GER.NOM.MSG
'Those monks … should perform the uposhatha ceremony.' (Pali, Mahāvagga II:11,1; Peterson 1998:97)

In (1.7 – 1.8), indexation is done by a suffix of the verb form. This is the most common strategy one across the data. But in some Central Iranian languages, such as

Gorani and Sivandi, indexation can also be achieved with a bound clitic pronoun, which can attach to constituent other than the verb form, and often attaches to the last preverbal phrase. This indexing pattern is often limited to non-monovalent verbs in the past tense. The factors that constrain the actual position of the clitic pronoun are not well understood, and seem to vary across the Central Iranian languages (see discussions in MacKenzie 1961, Lecoq 1979:40-42, Stilo 2004, Haig 2008:288-301, Mahmoudveysi et al. 2012:36-39). This indexing pattern is illustrated by (1.9a-b), where the clitic pronoun is attached to the last phrase before the verb, and indexes the A argument. In both examples, the A argument is also expressed in the clause by a NP or a free pronoun.

(1.9a) *tū* *ařā tʃe* *īma=t* *kyās* *ařā* *īnā*
2SG for what 1PL=2SG send.PST to here
'What did you send us here for?' (Gorani, Mahmoudveysi et al. 2012:39)

(1.9b) *espe=y* *siye=y* *harakat=eš* *kerd*
dog=LNK black=DEF movement=3SG do.PST
'The black dog moved away.' (Sivandi, Lecoq 1979:139)

Another axis of variation in terms of argument indexation across Indo-Iranian verbal clause constructions is its very existence. In some clause constructions across the family, there is no apparent argument indexation at all. This is illustrated by (1.10) from Middle Persian, where the verb is in a participle form which expresses necessitive

modality. Despite its obvious nominal origins (see brief review in Shirtz 2016), this form functions as the only verbal predicate in the clause in (1.10):

(1.10) *ā=š* *stōš* *ne* *yaz-išn*
then=3SG stoš.ceremony NEG celebrate.PRS-PTCP
'One should not perform the Stosh ceremony for him.' (Middle Persian, RAF 142A)

Constructions which do not exhibit any argument indexation are found in many languages across Indo-Iranian. One language in the data, Nagamese, has completely lost all verbal indexation constructions, and Nagamese verbs seem to be "stuck" in a form which corresponds to the Bengali / Assamese 3SG form. Thus, the verbs in (1.11a-b) below express TAM categories only:

(1.11a) *Ami apuni-khan nimite ekta khusi khobor an-ise*
1SG 2SG-PL for one good news bring-PRS
'I bring you good news' (Nagamese, Luke 2:10)

(1.11b) *tinihoile isor=bi* *apuni=ke maph* *kor-ibo*
and God=TOP 2SG=OBJ mercy do-FUT
'and God will forgive you' (Nagamese, Luke 6:37)

These axes of variation are not limited to crosslinguistic variation, but also figure in intralinguistic variation. Consider examples (1.7b) and (1.10) from Middle Persian, where in the former the verb indexes the S/A argument in number and person and in the latter it does not show any indexation. Examples (1.7a) and (1.8), from Pali, also vary in

their indexation properties; such intralinguistic variation in indexation is common across Indo-Aryan. Furthermore, in Gorani and Sivandi, indexation is also achieved by a set of verbal suffixes, cognate to the Middle Persian suffix in (1.7b), when the verb is in the present tense. In fact, the only language across the sixteen analyzed here in which no intralinguistic variation in indexation was found is Nagamese, which exhibits no indexing constructions at all.

1.3.3 Noun Flagging

The flagging of nouns, by case markers, adpositions, or a combination of both, is perhaps the most varied structural coding means in Indo-Iranian. Throughout the documented history of Indo-Iranian, there was a shift in the major flagging means used across the family, from the old Indo-European system of eight morphological case forms, to flagging mostly by adpositions or a combination of relics of the old case system together with innovative adpositions. In this section, I will briefly demonstrate some of the types of noun flagging found in the data.

The term "case" has been used in functional and typological linguistics in many distinct ways, which do not always end up referring to the same kinds of constructions. Some use the term "case" as a purely structural coding means (in the terminology used here, "flagging"), while others use it as a functional term for semantic roles (at least since Fillmore 1967). These two senses of the term are often found side by side in descriptions of Indo-Iranian languages, which can lead to some confusion.

For example, Kachru (2006:49, boldface mine) argues that Hindi makes a three-way distinction between "direct, oblique and vocative **case** forms. All other **cases** are

indicated by postpositions." This statement uses both distinct senses of "case", that is, as a morphological form class and for semantic roles. Further, an uneven application of the two senses can result in mismatches in the number of different cases reported for one and the same language, leading to possible confusion for readers. For example, Neukom and Patnaik (2003) report six different "cases" for Standard Oriya: Nominative, Oblique (base), Dative, Genitive, Ablative and Locative (2003:47), while Ray (2003:454) reports only three "cases": Nominative, Genitive and Objective. These differences are the result of different definitions of case.

In order to avoid such confusion here, I will reserve the term "case" for Old Indo-Iranian nominal case forms and cognates of those forms in Middle and Modern Indo-Iranian languages. Consider the flagging of the recipient participant in examples (1.12a-b) below, from Sanskrit and Sivandi, respectively. In (1.12a), from Sanskrit, the recipient is expressed by *abhimanyave* 'to Abhimanyu', a NP flagged by the Dative case form. In (1.12b), from Sivandi, the recipient is flagged by the preposition *ba* 'to, for' and is expressed by the PP *ba ossad* 'to the master'. In Sivandi, the recipient can also be expressed by an unflagged NP, as illustrated by (1.12c).[15]

(1.12a) *duhitaraṃ* *dattvā* *tām* *abhimanyave*
daughter.ACC give.ABS DEM.ACC.FSG Abhimanyu.DAT
'He gave his daughter to Abhimanyu.' (Sanskrit, Mahabharata 4.67.19)

[15] Unflagged R arguments in Sivandi are postverbal in all examples found in the texts used for this study. When Recipients in Sivandi are flagged by *ba* 'to, for', they are either preverbal or postverbal.

(1.12b) *de=yeš* *ba ossād*
give.PST=3SG to master
'He gave them to the master.' (Sivandi, Lecoq 1979:86)

(1.12c) *ow=rā* *be-de-y* *vi*
water=OBJ IRR-give.PRS-1SG willow.tree
'I shall give water to the willow tree.' (Sivandi, Lecoq 1979:85)

The reduction in the number of case forms distinguished in Indo-Iranian languages is illustrated by tables 1.1 – 1.4 below. Old Indo-Iranian languages, such as Sanskrit and Avestan, distinguished eight case forms in several different lexically-specified noun declensions (e.g., Masica 1991:230-231, Cardona 2003, Skærvø 2009). Table 1.1 shows the case forms of Sanskrit *deva* 'God' and Avestan *ap* 'water' in the singular.

Table 1.1: Case inflection of singular nouns in Sanskrit and Avestan

	Sanskrit *deva* 'God' (Cardona 2003: 143)	Avestan *ap* 'water' (Skærvø 2009:79)
Nominative	*dev-as*	*āf-ſ*
Genitive	*dev-asya*	*ap-ah*
Accusative	*dev-am*	*āp-am*
Dative	*dev-āya*	*ap-ai*
Ablative	*dev-āt*	*ap-at*
Instrumental	*dev-ena*	*ap-ā*
Locative	*dev-e*	*ap-i*
Vocative	*dev-a*	*āf-ſ*

Case syncretism had already begun in Old Persian, where the Genitive and the Dative cases merged, as did the Ablative and the Locative case, resulting in a system with six distinct case forms. A similar reduction can be identified in many Middle Indo-Aryan languages (see Oberlies 2003).

The trend of reduction continued in Middle Iranian, and in Sogdian, a Central Middle Iranian language, lexical nouns can be divided in two sets of stems, traditionally called "heavy" and "light" stems. The first set, "heavy" stems, distinguishes only two case forms, usually referred to as the Direct case and the Oblique case: for example, *mēθ* 'day' is the Direct case singular form, and *mēθī* is the Oblique case singular form. "Light" stems, in contrast, distinguish six case forms, shown in Table 1.2. The rise of these two distinct sets of nouns is usually attributed to changes in the placement of stress within the word (see Sims-Williams 1982, 1984, 1990; Yoshida 2009).[16]

As far as I can tell, no single Sogdian noun actually distinguishes all six case forms, and the analysis of a nominal system with six case forms (with "light" stems) can be argued for only by comparing different noun stems, like the three in Table 1.2. The different declensions found with Sogdian "light" stems are not attested in the plural, where a more agglutinative morphology developed.[17]

Table 1.2: Sogdian Light Stems (based on Yoshida 2009:288)

	Singular			Plural	
	ram- 'people'	*βaɣn-* 'temple'	*wan-* 'tree'	*ram-* 'people'	*Wan-* 'tree'
Nom.	*ram-i*	*βaɣn-u*	*wan-a*	*ram-t-a*	*wan-t-a*
Acc.	*ram-u*	*βaɣn-u*	*wan-a*	*ram-t-a*	*wan-t-a*
Gen.-Dat	*ram-e*	*βaɣn-e*	*wan-ya*	*ram-t-ya*	*wan-t-ya*
Loc.	*ram-a*	*βaɣn-ya*	*wan-ya*	*ram-t-ya*	*wan-t-ya*
Inst-Abl.	*ram-a*	*βaɣn-a*	*wan-e*	*ram-t-e*	*wan-t-e*
Voc.	*ram-a*	N/A	*wan-e*	*ram-t-e*	*wan-t-e*

[16] Sims-Williams 1982 shows that it is difficult to equate the Direct and Oblique cases that are found with "heavy" stems with the group of cases found with "light" stems.
[17] The Sogdian "numerative case", which flags nouns of the "light" stem set before numbers, is not included here.

In Early Middle Persian, a Western Middle Iranian language, the old case system had almost completely collapsed. Most nouns distinguished only two case forms, traditionally also called the Direct and the Oblique cases, as shown in Table 1.3 below. By Late Middle Persian, the variety analyzed here, even those distinctions had disappeared, and the Oblique Plural marker *–ān* functioned as a general plural marker (see Skjærvø 2009:205):

Table 1.3: Early Middle Persian case declension (Skjærvø 2009:205)

		mard 'man'	*pid* 'father'
Singular	Direct	*mard*	*pid*
	Oblique	*mard*	*pid-ar*
Plural	Direct	*mard*	*pid-ar*
	Oblique	*mard-ān*	*pid-ar-ān*

The state of the retained forms of the old case system in many modern Indo-Iranian languages resembles the situation found with Sogdian "heavy" stems, where nouns distinguish between only two case forms, again usually referred to as the Direct and the Oblique. This is shown by the Hindi noun declension in Table 1.4 below. Nouns in many modern Indo-Iranian languages (e.g., Early Judeo-Persian, Nagamese, Kupia) do not distinguish any case forms.

Table 1.4 Hindi noun case declension (based on Montaut 2012:102)

	Singular				Plural			
	Masculine		Feminine		Masculine		Feminine	
	'boy'	'house'	'girl'	'night'	'boy'	'house'	'girl'	'night'
Direct	*laṛk-ā*	*ghar*	*laṛk-ī*	*rāt*	*laṛk-e*	*ghar*	*laṛk-ā̃*	*rāt-ẽ*
Oblique	*laṛk-e*	*ghar*	*laṛk-ī*	*rāt*	*laṛk-õ*	*ghar-õ*	*laṛk-iyõ*	*rāt-õ*

Two-term case systems, composed of a Direct and Oblique cases, are quite common across Middle and Modern Indo-Iranian languages. The functions of these two case forms are crosslinguistically somewhat varied, but usually the Direct case form flags S arguments, and the Oblique case form flags NPs governed by some adposition. Further, in accusative clause constructions, the Direct case form often flags S/A arguments and the Oblique case form often flags P arguments; while in ergative clause constructions, the Oblique case form flags A arguments and the Direct case form flags S/P arguments. (See Arkadiev 2009 for a general discussion of the typology of two-term case systems, and Stilo 2009 for a discussion of two-term case systems in Iranian.)

Along with the decrease in the number of case forms, there seems to have been a rise in the overall number of innovative adpositions and their use in flagging core arguments and oblique objects. Some of this usage has already been demonstrated above, e.g., in example (1.12b) from Sivandi, where the recipient was flagged by *ba* 'to, for', and in (1.12c), where the theme argument is flagged by *=rā*.

The use of innovative adpositional elements to flag core arguments and oblique objects is further illustrated in (1.13a-c) below. In (1.13a), repeated from (1.1a) above, the A argument is flagged by *=ne* in the Hindi Ergative construction. In (1.13b), from Kupia, the P argument is flagged by *=ka*, and in (1.13c), from Middle Persian, an Oblique Object is flagged by *az* 'from':

(1.13a) *tum=ne kya kiyā?*
2SG=ERG what do.PRF.FSG
'What did you do?' (Hindi, Premchand 2017[1936]:5.61)

(1.13b) *wa:gu=ka dek-a ker-a*
tiger=OBJ see-NF do-NF
'I saw the tiger…' (Kupia, Christmas & Christmas 1973a:39)

(1.13c) *u=š az awēšān pursīd*
and=3SG from 3PL ask.PST.3SG
'He asked them.' (Middle Persian, DK6 D5)

In many, if not most, modern Indo-Iranian languages, the flagging of P arguments by an adposition in a differential object marking system is limited to P arguments that are higher in specificity and identifiability; these arguments are often also animate. Other P arguments are expressed by unflagged NPs. Compare, for example, (1.13b) above to (1.14) below, where the participant seen is general, low in specificity, and inanimate, and hence not flagged by an adposition. The actual conditions which constrain the employment of object markers are slightly different across the family (see, for example, Bossong 1985 for an overview of DOM in Iranian).

(1.14) *mu:rtumu dek-umde*
horoscope see-FUT.1PL
'We'll look at the horoscope.' (Kupia, Christmas & Christmas 1973a:6)

With personal pronouns one can discern a similar, albeit slower, process of loss of the old Proto-Indo-European system of eight distinct case forms. Across the Iranian branch, however, two distinct sets of clitic pronouns are found, usually one set per language, which are cognates of either the Old Persian Accusative Clitic Pronouns or the Old Persian Genitive-Dative Clitic Pronouns (see Korn 2009 for a detailed analysis).

Table 1.5 sets out the Middle Persian clitic pronouns, cognates of the Old Persian Genitive-Dative Clitic Pronouns (based on Sims-Williams 1981, Korn 2009, Durkin-Meisterernst 2014:209).

Table 1.5: Middle Persian Clitic Pronouns

	1st	2nd	3rd
Singular	*=(u)m*	*=(u)t, =(u)d*	*=(i)š*
Plural	*=(i)mān*	*=(i)tān, =(i)dān*	*=(i)šān*

The grammatical functions of these clitic pronouns vary across Iranian. In Middle Persian, for example, the clitic pronouns express P arguments with present tense verbs, A arguments with past tense verbs, S/A arguments in the "need" auxiliary construction, possessors in predicative possession constructions with "be/become", and some non-core functions (see overview in Haig 2008:105-131, Korn 2009, Shirtz 2016). In Gorani, for example, the cognate of the Middle Persian clitic pronoun is used in a different array of functions, and has grammaticalized as an indexing marker (see examples (1.9a-b) above).

1.3.4 Argument Structure

Finally, this subsection briefly presents Croft's model of argument structure constructions (2001:203-233), which is adopted in this study. Croft analyzes argument structure constructions not in terms of grammatical relations such as "subject" or "object", but instead in terms of symbolic, part/whole relationships. The main motivation for adopting Croft's model of argument structure is that it can be readily employed to identify whether a specific construction belongs to a given construction type.

The construction type in (1.1a-b) above, repeated below as (1.15a-b), was defined above as "a bivalent clause construction where the main verbal predicate is a cognate 'do/make' verb, and in which one argument, coded as an A argument, expresses the doer microrole and another argument, coded as a P argument, expresses an object the doer creates or a situation the doer is involved in." This definition can be easily interpreted using a part/whole model of argument structure. The semantic, or functional, pole of the construction type includes the two participants and their semantic microroles (e.g., doer and created element), and a "do/make"-type event. The syntactic, or formal, pole includes information about the coding of these participants in terms of comparative concepts (i.e., A and P arguments), and the lexical identity of the verb (i.e., a cognate "do/make" verb).

(1.15a) *tum=ne kya kiyā?*
2SG=ERG what do.PRF.FSG
'What did you do?' (Hindi, Premchand 2017[1936]:5.61)

(1.15b) *guru kicco ker-l-o?*
teacher what do-PST-MSG
'What did the teacher do?' (Kupia, Christmas & Christmas 1973a:75)

This approach to argument structure constructions is especially profitable for analyzing N-V complex predicate construction types in which one of the nominal constituents encodes the situation expressed by the entire clause. Such construction types are described in more details in Chapters IV and VIII, but for now, they are illustrated by (1.16a-b), where "do/make" is accompanied by a noun encoding the situation expressed

by the clause. In (1.16a) this noun is *judo* ‘fight’, and in (1.16b) *pašn* ‘contract, agreement’.

(1.16a) *bak ləgot judo kor-i bak mar-ise.*
tiger COM fight do-NF tiger die-PST
‘He fought with the tiger and killed it.’ (Nagamese, Sreedhar 1985:194)

(1.16b) *mard-ēw ke abāg zan ī xwēš pašn kun-ēd*
man-INDEF REL with woman LNK REFL contract do.PRS-3SG
‘a man who makes a contract with his wife’ (Middle Persian, RAF Q62)

In Chapters IV and VIII, this construction type is defined as a clause in which “a cognate ‘do/make’ verb is accompanied by a doer participant, coded as an A argument; an affected participant, coded by an NP flagged by a comitative marker; and a deed element, encoding the type of situation expressed by the clause.” On the semantic, or functional, pole of this construction one finds the doer and affected participants, which are linked, respectively, to the A argument and the comitative oblique object on the syntactic, or formal, pole. The type of event expressed by the clause is linked to the deed element, and the verb encoded verbal categories and perhaps also voice (see Korn 2013).

1.4 Summary

This study investigates the functional and structural diversification in the usage of three cognate verbs, “do/make”, “be/become”, and “give”, across a sample of sixteen Indo-Iranian languages. This chapter has presented the four sets of questions posed in the study, and the theoretical orientation of the study. This chapter has also presented the

main method that will be employed for crosslinguistic comparison, which is based on the definition of various construction types using "strategy-type" comparative concepts. Further, this chapter has also sketched the main mechanism of morphosyntactic change assumed in this study, and the way they interact with a constructional model of grammar. Finally, this chapter has given an overview of the different structural coding means which are often employed in argument structure construction across Indo-Iranian.

The rest of this study is set out as follows: Chapter II describes and motivates the type of data used in this study, i.e., published naturalistic texts accompanied by an authoritative translation. It then briefly describes the data sources used in this study and the languages sampled. It ends with a discussion of the notions of "doculect" and "languid", terms recently introduced in Cysouw and Good (2013) but implicitly and explicitly assumed in many descriptive works.

Chapters III – VI answer the first question posed in this study: what are the attested uses of "do/make", "be/become", and "give" across the sample of sixteen Indo-Iranian languages that is analyzed here? Chapter III re-introduces the notion of construction type, and illustrates it by several distinct construction types. It then provides an overview of the construction types attested with the verbs "do/make", "be/become", and "give". Chapters IV, V, and VI analyze in more detail the different construction types in which "do/make", "be/become", and "give" are attested in the data, respectively. Each of these three chapters begins with an analysis of the construction types in which these verbs function as main verbal predicates, including construction types which can be described as N-V complex predicates, and construction types expressing different nominal predication functional domains. Each of the chapters then describes the different

construction types in which the verb functions as an auxiliary, or at least is accompanied by another verbal predicate.

Chapter VII analyzes the use of "do/make" as the main verbal predicate in N-V complex predicate constructions, raising the question which synchronic types of argument structure constructions are used in N-V complex predicates, and whether there is an overall aggregate correlation between the type of situation expressed by the N-V complex predicate and the argument structure of the construction type. It then determines the diachronic processes by which the different attested types of N-V complex predicates with "do/make" evolved.

Chapter VIII analyzes the patterns in which different nominal predication functional domains are co-expressed by the same configurations of structural coding means. It begins with the observation, already suggested in Chapter V, that "be/become" verbs are used together with the same configuration of coding means in clauses expressing varying functional domains of nominal predication. Chapter VIII then presents a method for visualization and crosslinguistic comparison of these co-expression patterns, based on bipartite (bimodal) networks. Chapter VIII uses this method to compare (a) the co-expression of the predicate locative and core nominal predication (i.e., equation, predicate attribute, and proper inclusion); (b) the co-expression of predicative possession and the predicate locative, and of predicative possession and the existential; and, finally, (c) the co-expression of the predicate locative and the existential.

Chapter IX poses several nested questions about the aggregate crosslinguistic patterns in which "do/make", "be/become", and "give" are used. The first of these is whether the aggregate usage of these three verbs in closely related languages is more

similar than it is between? more distantly related languages. After demonstrating that the answer to this question is yes, this chapter examines the scope of the subfamilies within which these similarities can be identified, and the diachronic processes which drive these differences. This chapter uses Neighbor-Net (Bryant and Moulton 2004), a distance based method used to compare and visualized similarity between taxa. Chapter X summarizes this study and charts some future avenues.

CHAPTER II

THE DATA

This chapter describes and motivates the use of published naturalistic texts as the main data sources for this dissertation. The use of naturalistic texts as the primary data source is becoming more and more popular in comparative and typological studies, but is still rather uncommon. This section will begin with a brief explanation of the motivation behind this choice. This will be followed by a description of the languages sampled for this dissertation and the texts chosen to represent them. The section concludes with a short discussion of the methodological issues of varying textual genres. This chapter ends with a short description of the term "doculect", introduced in Cysouw & Good 2013, which is particularly appropriate for the current project, as well as for similar projects.

2.1 Data Sources

The Indo-Iranian language family is the largest sub-branch of the Indo-European phylum. This statement is true both in terms of number of attested languages (both extinct and living) and in terms of the number of speakers.[18] *Glottolog*, for example, identifies 584 Indo-European languages, and of these 318 are identified as Indo-Iranian (Hammarstöm et al. 2017). Similar figures are found in *Ethnologue* (Lewis et al. 2009).

Genealogically, the Indo-Iranian language family is divided into three main branches: Iranian, Indo-Aryan, and Nuristani. Glottolog identifies 95 languages as Iranian, 217 as Indo-Aryan, and six languages as Nuristani. The data for this study comes

[18] This statement is true also if one considers the Indo-Aryan sub-branch on its own.

from sixteen languages divided between the Iranian and the Indo-Aryan branches. The main reason behind the exclusion of the Nuristani languages from this dissertation is the lack of accessible and suitable data.

Historically, Indo-Iranian languages have been documented since the second millennium BCE, the earliest known attested forms being Vedic Sanskrit on the Indo-Aryan side and Avestan on the Iranian side. Documentation of Indo-Iranian languages has been more or less uninterrupted since that time. Geographically, Indo-Iranian languages are spoken across a vast area stretching from Nagaland in the northeast of India and the island of Sri Lanka to the southeast of India, to central Turkey and the Caucasus mountains in the west. In the post-colonial period, communities of Indo-Iranian speakers are found also elsewhere (e.g., the so-called Hindi varieties of Fiji and Guyana, Romani and Domari speakers in the Middle East and Europe).

The historical depth of extant documentation, together with the vast geographic span of these languages, makes it difficult to encompass all the grammatical variation found across Indo-Iranian[19] in a single work, even when the features are limited to the grammar associated with just three cognate verbs. As a result, there is much potential for replication and refinement of the analyses given here with an entirely different set of languages, or an entirely different set of texts in the same languages.

[19] Consider, for example, the diversity found in alignment patterns across different branches of Indo-Iranian (e.g., Payne 1980, Deo & Sharma 2006, Haig 2008, Verbeke 2013, *inter alia*).

2.1.1 Type of Data

As mentioned above, the main data source for this dissertation is published textual documentation of naturalistic language use. These texts are of various types, and include texts published as a part of a language documentation project, texts written for a native-speaker audience (i.e., literary prose), or, in the case of extinct languages, published textual editions of manuscripts. The patterns of usage of the three cognate verbs analyzed here ("do/make", "be/become", and "give") are identified by analyzing all occurrences of the three verbs in these textual data. Most examples given in this dissertation are taken from these texts. Some of the construction types in which these verbs are attested, especially auxiliary constructions, have been thoroughly analyzed in the linguistic literature (i.e., grammars, grammatical sketches, and dedicated papers), and the analysis of these usages makes reference to this literature and the examples used there.

In contrast to the approach in this study, the main data source used in many other typological and comparative studies is published descriptions of different languages, either as grammars or as studies focused on specific phenomena. Typological databases such as WALS (Dryer et al. 2013), for example, make almost exclusive use of published grammars and grammatical analyses. In recent years, however, several studies have used naturalistic texts as their main source of data, and this trend seems to be more and more popular. This includes studies in the Preferred Argument Structure theory such as Du Bois 1987, the studies published in Du Bois et al. 2003, and studies arguing against them, e.g., Haig & Schnell 2016. They also include comparative studies like Verkerk (2014), for example, who used translations of *Alice in Wonderland*, *Through the Looking Glass*, and the *Alchemist* into 20 modern Indo-European languages as the main source of data

for her analysis and for phylogenetic reconstruction of the patterns of the expression of motion across Indo-European. Similarly, Hartmann et al. (2014), Cysouw (2014) and Cysouw & Wälchli (2012) use Bible translations or translations of Harry Potter novels as their main data source in several recent projects.

The motivation for using naturalistic texts rather then grammatical descriptions in this study is double. First, the data required to answer most of the questions posed here is not available in published grammatical descriptions. While cognate "do/make", "be/become", and "give" verb, which are the topic of this dissertation, do figure here and there in descriptions of Indo-Iranian languages and analyses of various grammatical phenomena, such analyses are mostly limited to the more grammatical functions of these verbs such as their usage in auxiliary constructions. As a result, such analyses often leave out some "mundane" or theoretically less interesting usages of these verbs. Furthermore, many analyses of some theoretically central phenomena in Indo-Iranian, such as complex predicate constructions, do not necessarily limit their scope to the use of "do/make", "be/become", and "give" in their constructions of interest. Thus, those analyses cannot be used as the sole source for the grammatical behavior of these three verbs in such constructions, and textual data is needed.

Finally, much of the literature dedicated to the study of specific constructions in Indo-Iranian languages concentrates on a rather small group of (relatively) well-researched languages spoken by a large number of speakers, such as New Persian or Hindi.[20] Thus, analyses and data about the use of these three verbs in specific functions in

[20] This is not to say that there are no remaining problems in the analysis of different phenomena in these languages. The vast, and relatively recent literature on complex predicates in New Persian demonstrates that our understanding of some phenomena is far from complete.

other, often minority, languages, is rare in the literature.[21] In order to adequately characterize the diversity in the usage patterns associated with these three verbs and chart the changes in their grammar across the Indo-Iranian language family, one needs to use primary texts from a variety of languages as the main data source, and to use analytic publication such as grammars in a secondary way, to complement it.

The second reason for using naturalistic usage data for this project has to do with the vexed definition of what a specific language is, and the constant intralinguistic variation (dialectal, social, or otherwise). Many Indo-Iranian languages that would "fit" under a single unique language name (such as "Middle Persian" or "Gujarati") are further dividable into different varieties with (sometimes) considerable grammatical differences. These varieties can be the result of diglosia (e.g., Jeremeias 1984 for Persian, Krishnamurti et al. 1986 for Indo-Aryan), dialectal variation, a single ethnic group which members do not share a single language (e.g., Kurdish; see Haig & Öpengin 2014) and so on. In some cases, variation is also connected to the expression of grammatical categories, and is not limited to languages with larger numbers of speakers such as Persian, Kurdish, or Hindi.

Stilo 2004, for example, identifies inter-generational differences in the coding of core arguments in Gazi, a Central Iranian language spoken in the city of Gaz. These differences may be also correlated with differences in education levels and exposure to New Persian. Dialectal variation seems to have been acknowledged in Hook's 1995 study

[21] Recently, Trudgill 2011 suggests a link between linguistic structure on the one hand, and on the other, differences in social structure of the speech community and the status of (or attitude towards) the language in the community. Limiting the range of languages sampled here to written, standard(ized) languages spoken by large communities of L2 speakers might have led to a different result.

of Gujarati compound or conjunct verbs (a specific type of Indo-Aryan auxiliary construction, see Chapter VI for discussion), as the results of his study differ from those of Christian 1987, where the source used for the current project is published.

The effect of intralinguistic variation on linguistic analyses might also underlie the differences in acceptability judgments of some grammatical constructions. For example, Ghomeshi & Massam 1994 argue for the absence of a double object constructions from New Persian. There, this claim was made in order to reject a particular analysis of complex predicates (see Chapter VII for further details). As has been pointed out in Müller 2010 and elsewhere (e.g., Lazard 1994, 2012), double object constructions actually do occur in New Persian. The primary data source of Ghomeshi & Massam 1994 on the one hand, and of Müller 2010 on the other hand, is acceptability judgments. As neither of them mentions the linguistic background of their consultants (or their number, or the specific stimuli used and elicitation technique), these differences might be attributable to dialectal differences or other factors correlated with intralanguage variation.

Using published naturalistic texts as the main data source makes it possible to directly test and compare different analyses. For example, one could potentially apply the analysis of Hook's 1995 Gujarati Compound / Conjunct verb constructions to the data in Christian 1987 and vice versa, and thus directly compare both analyses. Admittedly, this can also be done if a linguist uses acceptability judgments or elicited sentences as her main data source, but it requires a more detailed account of the linguistic background of consultants, the elicitation method and the stimuli used than is usually given.

2.1.2 The Sample

As mentioned above, it is virtually impossible to encompass the entire geographic breath and historical depth of the Indo-Iranian language family in a sample of sixteen languages. In creating the sample for this dissertation, I attempted to include representatives of the three main historical stages of Indo-Iranian (traditionally called Old, Middle, and New), as well as (at least) several major branches of both Indo-Aryan and Iranian. For Indo-Aryan, the sample contains data representing Northern, Central, Western and Eastern Indo-Aryan languages. For Iranian the sample has data representing Central and Western Iranian languages. I also tried to bias the sample towards minority languages.

The languages sampled here, their classification and ISO codes (where available), and the textual sources used for the current project are described in Table 2.1. This table is followed by a short description of each language and the data source(s) used.

As noted earlier, the data sample used for this project cannot cover the historical, geographic, and genealogical breath and depth of Indo-Iranian. Certain languages of interest had to be excluded from this sample, either because of the lack of appropriate documentary material or because of space and time considerations. Nevertheless, while lacunae in the coverage of this sample are unavoidable, they do allow for testing and refinement of the results presented in this dissertation using a different set of data.

Table 2.1: The data sample.

Language name	**Classification**[22]	**ISO code**	**Primary sources**
Sanskrit	Old Indo-Aryan	SAN	Mahabharata, book 4 (Buitenen 1973, Garbutt 2006)
Bactrian	Middle Central Iranian	XCD	Bactrian documents and letters (Sims Willaims 2000, 2007)
Sogdian	Middle Central Iranian	SOG	MacKenzie 1979, Sogdian Buddhist Documents
Late Zoroastrian Middle Persian	Middle Western Iranian	PAL	Portions of three texts from the Middle Persian Dictionary Project: the tale of righteous Wiraz (Vahman 1988, narrative), DK6 (religious parables, Shaked 1979), Pahlavi Rivayat (religious discussions, Williams 1990)
Early Judeo Persian	Early Modern Western Iranian	N/A	Fifty pages from Gindin 2008 (exegesis of the book of Ezekiel).
Pali	Middle Indo-Aryan	PLI	20 Jataka (previous births of the Buddha) tales from the *suttapiṭaka*.
Hindi	Modern Indo-Aryan, Central	HIN	4 first episodes of Premchand's *Godaan* "A gift of a cow" (1936)
Kotia (Adivassi) Oriya	Modern Indo-Aryan, Eastern	ORT	Texts published in Gustafsson 1973
Kupia	Modern Indo-Aryan, Eastern	KEY	Texts in Christmas & Christmas 1973
Gujarati	Modern Indo-Aryan, Western	GUJ	Folk tales published in Christian 1987
Nagamese	Modern Indo-Aryan, Eastern	NAG	Nagamese Bible: Jonah, 6 chapters from Luke. Texts published in Sreedhar 1984
Palula	Modern Indo-Aryan, North-Western	PHL	Texts published in Liljegren & Haider 2015
Ziyarat Mazandarani	Modern Central Iranian	N/A	Texts published in Shokri et al 2013
Gorani	Modern Central Iranian	HAC[23]	Texts published in Mahmoudveysi et al. 2012
Sivandi	Modern Central Iranian	SIY	Texts published in Lecoq 1979
Darai	Modern Indo-Aryan, Eastern	DRY	Texts published in documentation project in Dhakal 2013

Apart from time, the crucial practical consideration in sampling the languages for this study was the availability of published textual material 10,000 to 15,000 words long,

[22] The classification of Iranian is give according to Korn 2016.
[23] This ISO term includes several other languages as well.

accompanied by an authoritative translation into English, French, or German. That is, for each text, I required a translation by a native speaker for texts written for native audience, a linguist collaborating with native speakers for texts published as a part of a documentation project, or, in the case of texts from an extinct language, a philologist preparing a critical edition of a text or a manuscript. The main reason behind this requirement is that the three verbs whose usage patterns are analyzed here are often used in idiomatic patterns and their usage is often metaphorically extended.

The next sixteen subsections present the languages sampled for this dissertation and the texts used as data. The presentation is meant to be short, almost telegraphic in nature. Further data about the different languages, however, can be found in the sources cited in the next sixteen subsections.

2.1.2.1 Sanskrit

Sanskrit (*sãskrta* 'purified') is an Old Indo-Aryan language. Its further classification into one of the Indo-Aryan sub branches is unclear, and it is usually considered a (direct) daughter language of Proto-Indo-Aryan. Masica 1991 and Cardona 2003, for example, as well as many others, do not offer further classification of Sanskrit. Sanskrit was spoken, or used, over most of the northern Indian sub-continent from the middle of the second millennium BCE. The extant Sanskrit corpus is vast, both in terms of the number of texts which are available to us and in terms of historical depth. The earliest attestations of Sanskrit are in the Vedas, and the language used in them is usually referred to as Vedic Sanskrit. Non-Vedic literature, which is often divided into Epic Sanskrit (including the Ramayana and the Mahabharata), Classical literature (including

both scientific and poetic works), and Buddhist and Jain literature, is documented also well past the Tenth Century C.E.

It should be noted that a large portion of the Sanskrit literature was written during a time period that followed the emergence of Middle Indo-Aryan vernaculars (often known as *prakrit* 'natural, ordinary'). To illustrate, consider that in many Classical Sanskrit plays, not all characters actually speak Sanskrit. While sages, kings, and army generals do speak Sanskrit, other characters, such as jesters and servants, speak a form of Middle Indo Aryan (often called *apabhramśa*). For example, in Kalidasa's famous play *abhijñānaśākuntalam* 'The recognition of Shakuntala', one of the pinnacles of Classical Sanskrit literature, Shakuntala herself speaks in a Middle Indo-Aryan language. The Middle Indo-Aryan lines were translated into Sanskrit and produced in a sort of an appendix to the play called *chaya* 'shadow'.

The Sanskrit data for this study is the fourth book of the *Mahābhārata*. The Mahabharata belongs to the Epic Sanskrit variety, and was mostly written in verse. The Sanskrit data is based on the editions available in the GRETIL corpus database (the Göttingen Register of Electronic Texts in Indian languages) available online, the Clay Library version of the text (Garbutt 2007), and the translation in Buitenen 1973.

2.1.2.2 Pali

Pali is a Middle Indo-Aryan language. Together with Sanskrit it is one of the major languages of Buddhist literature. Pali is usually not classified into one of the sub-branches of Indo-Aryan. However, Oberlies (2003) argues that when comparing Pali to

the language of Aśokan inscriptions from across South Asia, it is most similar to varieties found in Western and Northwestern India.

The majority of Pali texts, called the *tipiṭaka*, 'three baskets,' belong to the Theravada Buddhist canon. The three baskets consist of the *vinayapiṭaka* which deals with rules related to the Buddhist monk order; the *suttapiṭaka,* which contains the sermons of the Buddha as well as some miscellaneous texts from different Pali strata (Oberlies 2003); and the *abhidhammapiṭaka.*

The Pali data for this study is the *Jataka*, 'birth, previous birth', tales of the Buddha. It is a part of the the *khudakanikaya* collection of the *suttapitaka,* which consists of more or less miscellaneous texts. The edition used here is from the GRETIL corpus. The translation used here is mostly based Caldwell 1895.

2.1.2.3 Hindi

Hindi, a New Indo-Aryan language, is one of the most commonly spoken languages in the world today. *Ethnologue* (online version; Lewis et al. 2009) lists 260 million native Hindi speakers (mostly in India), and 120 million non-native speakers (in India). It is spoken by large communities across Western and Central Northern India (in an area sometimes referred to as the "Hindi Belt"; see, e.g., Masica 1991). It is usually classified as a Western or Central-Western Indo-Aryan language (see appendix 2 in Masica 1991 for details).

There are many different varieties and dialects of Hindi, as well as many recognized registers (see Shapiro 2003 for details and references). These varieties include both geographical and social dialects, as well as different historical stages of Hindi. Just

as in the case of Sanskrit, this variation makes it difficult to chose a single text to represent this language in the data for this study. As such, I do not consider the text chosen (or any single text, for that matter) to represent Hindi as a whole - this would be, I believe, grossly inaccurate, and perhaps impossible. Instead, it represents one instance of Hindi use, intended for a native audience.

The Hindi data consist of the first four chapters of Munshi Premchand's novel *Godaan* "A cow-gift". It is a literary text, originally published in 1936, and translated to English several times. It is difficult to determine whether the language of *Godaan* represents a specific dialect or variety other than that of Standard Hindi. Munshi Premchand, however, was born in a village near Varanasi and died in Varanasi itself, shortly after *Godaan* was published.

2.1.2.4 Gujarati

Gujarati is a New Western Indo-Aryan language. It is the official language of the modern Indian state of Gujarat in Western India. Gujarati speaking communities, however, can also be found in Mumbai and in Rajasthan as well as elsewhere throughout (especially Western) India and in some regions of Pakistan.

The Gujarati data for this study come from texts published in Christian 1987. These texts were collected in Gujarati speaking villages, and differ from written varieties of Gujarati in several respects, such as the function of "conjunct verb" construction (see Hook 1995). The texts published in Christian 1987 are all folktales, published with glosses and translated into English. Examples from these texts presented here have been reglossed to conform to the conventions used in this dissertation.

2.1.2.5 Kotia Oriya

Kotia Oriya, sometimes referred to as Adivassi Oriya, is a New Eastern Indo-Aryan language. It can be thought of as a variety of Oriya, one of the major Eastern Indo-Aryan languages and the official language of the Indian state of Oḍisha. Kotia Oriya is spoken in the northern part of the state of Andhra Pradesh (south of Oḍisha), in Vishakhapatnam district. There are 200,000 (Lewis et al. 2009, Gustafsson 1973b) native speakers of Kotia Oriya.

The Kotia Oriya data comes from texts collected as a part of a documentation project by Uwe Gustafsson, and published together with a preliminary grammatical analysis in Gustafsson 1973a,b. These texts include short narrative texts and descriptions of culturally significant events. The texts are all published glossed and translated. In the presentation of data from Kotia Oriya in this dissertation, I have reglossed the examples cited here, as the conventions for glossing used in Gustafsson 1973a are different from those used here.

2.1.2.6 Kupia

Kupia is a New Eastern Indo-Aryan language. Like Kotia Oriya, it is also spoken in the Vishakhapatnam district of the state of Andhra Pradesh. Speakers of Kupia are sometimes referred to as the Valmiki people. The Valmikis are a scheduled tribe of Andhra Pradesh (http://censusindia.gov.in/Tables_Published/SCST/ST%20Lists.pdf, consulted October 2016). According to *Ethnologue* (Lewis et al. 2009), there were under 7000 Kupia speakers in 2007, out of 79,000 Valmiki people.

Christmas & Christmas 1973b, as well as *Ethnologue*, report that in social activities many speakers of Kupia often use Telugu (Dravidian), the official language of the Indian states of Andhra Pradesh and Telangana. As a result, there are many easily identifiable influences of Telugu on Kupia, ranging from lexical borrowing (e.g., *dabbu* 'money'), to affix borrowing (e.g., the plural suffix *-lu*), to several clause pattern types either borrowed or heavily influenced by Telugu grammar.

The Kupia data for this study come from glossed and translated texts collected and published by Christmas & Christmas (1973a). Their publication of the texts is accompanied by an analysis of basic clause and sentence patterns in Kupia (Christmas & Christmas 1973b). The Kupia examples presented here also have been reglossed. The texts include short narratives and description of culturally significant events.

2.1.2.7 Darai

Darai is a New Indo-Aryan language spoken in Nepal. Its subclassification is uncertain (Dhakal 2013, Lewis et al. 2009), but it does contain some properties of Eastern Indo-Aryan languages. It is spoken by members of the Darai people across three districts of Nepal (Chitwan, Tanahun, and Nawalparasi, CBS 2002, cited by Dhakal 2013). Darai speakers have been in long standing contact with speakers of Tibeto-Burman languages of the area, and some grammatical patterns in Darai including verb indexing of the possessor (see Dhakal 2015) seem to have developed under the influence of from Tibeto-Burman (see DeLancey 2011).

The Darai data for this dissertation comes from glossed and translated Darai texts published in Dhakal 2013. These texts contain short narratives, explanations and descriptions of cultural activities, and an interview.

2.1.2.8 Nagamese

Nagamese is a New Eastern Indo-Aryan language, spoken in the Indian state of Nagaland in Northeast India. It is often considered to be a pidgin (e.g., Sreedhar 1985) or a creole language (e.g., Bakker et al. 2011), but it is not represented in *The Atlas of Pidgin and Creole Language Structure* (Michaelis et al. 2013) and exhibits either the lowest or close to the lowest number of creole grammatical features (depending on the specific feature list and definitions) tested by Bakker et al. 2011.

This dissertation treats Nagamese as an Eastern Indo-Aryan language, closely related to Assamese and Bengali,[24] which has greatly simplified certain aspects of its grammar, especially in morphology. Even if one objects to this classification, holding that Nagamese is actually a creole language and hence, somehow, "out" of the usual genealogical classification schemata, the inclusion of Nagamese is still justified by the fact that this dissertation deals with the grammatical behavior of cognates of Proto-Indo-Iranian "do/make", "be/become", and "give", and cognates of these three verbs are found in Nagamese. The social process that led to the creation of Nagamese as a distinct language does not figure in the definition of cognate.

[24] The exact genealogical relationship between Nagamese and other Eastern Indo-Aryan languages (especially Bengali, Sylheti, and Assamese) remains to be shown.

The Nagamese data come from two sources. The first source is the Nagamese translation of parts of the Bible: the book of Jonah and the first six chapters of the gospel of Luke. Apart from these, the other source is the short texts published in Sreedhar 1985. All these texts are narrative in nature.

2.1.2.9 Palula

Palula is a New Indo-Aryan language spoken in the Hindu-Kush region (Chitral district) of Pakistan. Liljegren (2016:13-20), considers Palula to be a member of what he calls the Hindu-Kush Indo-Aryan language group, and argues that it is closely related to Shina. Other sources, such as *Ethnologue*, consider Shina to be a member of the Dardic or Northwestern ("outer") sub-branch of Indo-Iranian (Masica 1991). Liljegren 2016 argues that there are about 10,000 speakers of Palula, with some monolingual (or almost monolingual) communities. This figure is also cited by *Ethnologue*.

The Palula data comes from glossed and translated Palula texts published in Liljegren and Haider 2015. These texts include mostly short narratives and descriptions of cultural events. A secondary source used here are the examples found in Liljegren 2016.

2.1.2.10 Sogdian

Sogdian is a Middle Central Iranian language. It was spoken throughout central Asia between the Third and Tenth centuries. Sogdian texts can be divided into several varieties, which can be classified by the community which produced them. These varieties differ in their grammar and the script used in the writing system. Three varieties

with a relatively substantial corpus are Manichean Sogdian, Christian Sogdian, and Buddhist Sogdian (for an overview see Yoshida 2009).

The Sogdian data for this study come from the Buddhist Sogdian texts published in MacKenzie 1976, based on manuscripts housed in the British library. These texts are published alongside an English translation, and an extensive glossary and concordance. They are mostly translations from Chinese and Sanskrit.

2.1.2.11 Bactrian

Bactrian is a Middle Central Iranian language. Prior to the recent publications of legal documents and letters (Sims-Williams 2000, 2007), published texts of any length in this language were rather scarce, and included mostly short inscriptions. Bactrian was one of the main languages of the Bactrian Empire in what is modern day Northern Afghanistan. For a short description of the history of Bactrian research, see Korn's 2015 review of Sims-Williams 2000 and 2007.

The Bactrian data for this study come from the texts published in Sims-Williams 2000 and 2007. These texts were published together with an English translation, and are accompanied by a grammatical sketch, a glossary, and a concordance. Bactrian examples presented in this dissertation were transliterated from the Bactrian alphabet (Greek alphabet plus one letter) and glossed using the glossary and grammatical sketch in Sims-Williams 2000, 2007.

2.1.2.12 Middle Persian

Middle Persian was a Middle Western Iranian language. It was in use between the third and the eleventh centuries, and alongside Aramaic, was one of the main languages of the Sasanian empire. There are (at least) three different varieties of Middle Persian, which can be roughly described as Inscriptional Middle Persian, or the Middle Persian found in royal inscriptions; Manichean Middle Persian, or the Middle Persian found in (mostly translated) Manichean texts; and Zoroastrian Middle Persian, sometimes also called "Book Pahlavi".

The texts used for the current project are all relatively late Zoroastrian texts, written in a writing system which derives from a contemporary Aramaic script. The cultural influence of Aramaic on the writing system was so great that about 800 Middle Persian words are written as Aramaic ideograms in Zoroastrian Middle Persian. These ideograms are often accompanied by Middle Persian, and not Aramaic, grammatical affixes. Zoroastrian Middle Persian has the largest surviving corpus of texts of all three varieties mentioned above, most of which was written originally in Middle Persian and not translated from other languages as the majority of Manichean texts were.

The Middle Persian data for this dissertation is composed of three texts. These are the first part of the sixth book of the Denkard (DK6, Shaked 1979), the first 20 pages of the book of righteous Wiraz (AWN, Vahman 1986), and the first 21 sections of the "Middle Persian miscellaneous" texts (PRDD, Williams 1990). Each of these texts belongs to a slightly different genre: DK6 is mostly composed of short parables and gnomic sayings and advice; AWN is a narrative text depicting the journey of Wiraz in heaven and hell; PRDD is composed of several different types of texts, including short

narratives and short descriptions of proper and desired behavior. In Chapter VII, where the emergence of complex predication in Middle Persian is described, these texts are supplemented by other late Middle Persian texts which are included in the online corpus of the Middle Persian Dictionary Project (MPDP).

2.1.2.13 Early Judeo-Persian

Early Judeo-Persian is a New Western Iranian language; it is a form of very early New Persian written in Jewish communities in 12th century Persia. A large number of Persian language texts from that period were written by members of the Zoroastrian and Jewish religious minorities. Texts in Early Judeo-Persian were written using the Hebrew script, and contain a large number of Hebrew and Aramaic loan words.

It is important not to confuse Early Judeo-Persian, which is a form of Early New Persian, with other Jewish Iranian languages spoken by Jewish communities in Iran, Central Asia, and the Caucasus. These languages are not necessarily daughter languages of Early Judeo-Persian, but are distinct languages spoken by (mostly multilingual) Jewish communities across the region, some of which belong to the Central branch of Iranian (e.g., Stilo 2007, Borjian 2014), while others belong to the Western branch but are not a form of New Persian (e.g., Bokhuri spoken by Jewish communities of Samarkand, and Juhuri spoken by some Jewish communities of the Caucasus).

The Early Judeo-Persian data for this study comes from the longest published text in Early Judeo-Persian: the exegesis of the book of Ezekiel (published in Gindin 2008) which contains a verse-by-verse translation of the book of Ezekiel into Early Judeo-

Persian, followed by an interpretation and a discussion. The translations were not themselves analyzed in this study.

2.1.2.14 Sivandi

Sivandi is a New Central Iranian language, spoken in the village of Sivand, about 50 miles north of Shiraz. Lecoq (1979) mentions about 4000 inhabitants of the village, but *Ethnologue* lists 7000 speakers for this language (giving only the year number 2000 as a reference). Whatever the number is, it seems clear that the language is endangered or at least highly threatened.

The Sivandi data come from the texts and analysis published in Lecoq 1979. The texts there are accompanied by a French translation, a glossary, and a short sketch of basic Sivandi grammar (mostly phonology and morphology). The examples presented here were glossed using Lecoq's glossary, morphological sketch, and translation, and are translated into English by me, in consultation with Lecoq's 1979 French translations.

2.1.2.15 Ziyarat Mazandarani

Ziyarati Mazandarani (henceforth Ziyarati or Ziyarat) is a New Central Iranian language spoken in the small village of Ziyarat in the Elborz mountains (Golestan province), near the southeast corner of the Caspian Sea, which has about 2000 inhabitants (Shokri et al. 2013). *Ethnologue* does not include a separate entry for Ziyarat, but lists over two million speakers for Mazandarani.

The Ziyarati data for this study come from the five texts published in Shokri et al. 2013, which are accompanied by translations into English and New Persian, a short

grammatical sketch and a glossary. Examples of Ziyarati presented in this dissertation were glossed in reliance on the grammatical sketch and the glossary. The texts contain sociolinguistic interviews, in which members of the Ziyarati community describe their lives, daily routines, and culturally significant events. For the most part, the interviews are conducted in Ziyarati, with the occasional slip into New Persian.

2.1.2.16 Gorani

Gorani is a New Central Iranian language, spoken in the village of Gawarju in Iran. Paul (2007) claims that there are about 180,000 speakers of Gorani, a number also cited by *Ethnologue*. The language is also known sometimes as Hawrami, or Avroami. There have been several publications of texts and glossaries of different varieties of this language.

The Gorani data used here come from the published Dobes documentation project in the village of Gawarju (Mahmoudveysi et al. 2012). The texts there include mostly folktales and sociolinguistic interviews. One of the texts is glossed and all the texts are accompanied by an English translation. Mahmoudveysi et al. 2012 also contains a glossary and a short grammatical sketch. Occasionally, speakers in the texts in Mahmoudveysi et al. 2012 slip from Gorani into Kurdish. Of course, these instances are not included here.

2.2 Different Texts – Different Genres?

As the sketches given in the previous section show, the different texts used in this study represent different textual genres. For example, the Bactrian data contains mostly

legal documents and letters, the data from early Judeo-Persian comes from a scholarly text (the exegesis of the Book of Ezekiel), and data for many of the modern languages are mostly composed of narratives and sociolinguistic interviews. Moreover, while most texts in the data are prose, the Sanskrit texts are in various types of verse.

The main problem that the range of genres poses to this study is the (very real) possibility that differences in the presence or absence of some usages identified here are motivated by differences in genre rather than differences in language. For example, the motivation for the "translational" or copula-like usage of "do/make" in Early Judeo-Persian (see Chapter IV for details) may have to do with the nature of the texts used here. It could be hypothesized that had texts of comparable genres in some other languages (e.g., Sivandi or Darai) been used, a comparable usage pattern of "do/make" might have been found in in them as well.

There is a clear tradeoff between the need to include historical data in the sample and a requirement for a unified textual genre across the data, and thus, more direct comparability. Given the current inventory of texts available in Indo-Iranian languages, a clash between these two requirements is unavoidable. Surviving texts in the older languages are sometimes limited in genre, style, and scope. Hence, if the genre had been kept as uniform as possible, it would not have been possible to include Bactrian or Early Judeo-Persian in the corpus, such that some stages of Indo-Iranian could not have been represented. I have chosen to include these languages, and to monitor the usage patterns unique to them for possible interactions with genre. While it is indeed possible, and, I believe, true, that there is interaction between genre and (the availability of) some usage

patterns of "do/make", "be/become", and "give", this is still an open question for future research.

2.3 Language (– "Languid") – Doculect

Kos kos pe pani bədle, car kos pe bani

"the water(shed) changes every kos (~2 miles), every four kos, the speech" (Old Hindi proverb, Dain 2003:47)

This short Old Hindi proverb demonstrates some folk knowledge of the linguistic variation of Northern India. A similar statement could easily apply to some areas where the languages of the Iranian branch of Indo-Iranian are spoken. As mentioned earlier in this section, one of the main reasons for the use of published textual documents for this study has to do with the grammatical variation found in what is often considered a single "language".

Many Indo-Iranian languages are spoken by multiple ethnic, social, and geographic groups of speakers and these may vary considerably in their grammatical patterns. This variation sometimes leads to differences in the usage patterns of certain grammatical constructions in naturalistic texts, or to differences in acceptability judgments by different speakers. Thus, I believe at least some, though definitely not all, disagreements in the analysis of various constructions in Indo-Iranian languages are the result of contrasting data sources different authors have used.

Variation across data sources is one of the reasons for the notion "doculect" in Cysouw & Good 2013. A "doculect" is defined there as a pair of a language name (such

as “Hindi” or “Gujarati”) together with a well defined reference for primary data (such as a publication of texts, a documentation project or a reference grammar). Acknowledging that the data for this study, and the data in many linguistic analysis projects, might involve not necessarily only different “languages” but also different doculects of a given language (“glossnyms” or “languids” in the system proposed in Cysouw & Good 2013) should assist in respecting, documenting and analyzing grammatical variation.

That different data sources which document distinct instances of language use might vary in their grammar despite being of “the same language” has been long acknowledged in some branches of linguists. Many philological works explicitly limit their linguistic analysis to one written source of the language (e.g., Barri 1977’s analysis of clause patterns in the language of Antiphon). Within Iranian, one can mention Tadesco’s 1921 analysis of the differences between Iranian manuscripts found in Turfan (Western China), which enabled the identification of two distinct Middle Iranian languages, Middle Persian and Parthian, which belong to different branches of Iranian. Recently, Josephson 2016 and Shirtz 2016, have analyzed Late Middle Persian and both argue that some of the phenomena they identify are limited, at least to some degree, to Late Middle Persian texts. Variation between different sources is also identified in this study, for the expression of Predicative Possession with “be/become” in Middle Persian (see Chapter V).

Thus, to delimit the scope of this study and the data used in it more precisely, this study describes and analyses the usage patterns of three cognate verbs (“do/make”, “be/become”, and “give”) across a set of sixteen of Indo-Iranian doculects, rather than entire languages. Therefore, every time a reference is made to language by its name (e.g.,

"Middle Persian" or "Hindi"), the name of the language should be understood to mean the doculect of that language as represented by the primary data source used here (cf. table 2). This does not mean that published analyses of similar phenomena in the languages of the sample (as well as other Indo-Iranian languages) are ignored in this study because these do not represent the "same doculect". However, I have kept in mind that any mismatches between a given analysis found in other works and those presented here may be due to the use of different sources.

CHAPTER III

CONSTRUCTION TYPES OF COGNATE VERBS

This chapter provides a more detailed description of construction types, already mentioned in Chapter I. It demonstrated the application of the term to several sets of examples, as a precursor to Chapters IV through VI of this study, which analyze the attested uses of the sets of cognate "do/make", "be/become", and "give" verbs in the sample of sixteen Indo-Iranian languages described in Chapter II. Chapter IV concentrates on constructions involving "do/make", Chapter V on "be/become", and Chapter VI on "give".

As noted in Chapter I, the language specific nature of grammatical categories and constructions poses a challenge to crosslinguistic comparison. The solution of many linguists interested in comparative (typological or historical) questions is to define "comparative concepts" (e.g., Hockett 1955, Stassen 1985:14, Haspelmath 2010, Croft 2016): concepts created by linguists for the purpose of crosslinguistic comparison, defined without any reference to language specific categories or constructions.

To compare the uses of "do/make", "be/become", and "give" across Indo-Iranian, this study examines them in terms of construction types, following Croft's definition of "strategy" type comparative concepts (2014:537; see also Chapter I). These can be thought of as bundles of comparative concepts which are a part of a single construction. For example, in Section 1.1, I define one (common) construction type as "a bivalent clause construction where the main verbal predicate is a cognate "do/make" verb, and in which one argument, coded as an A argument, expresses the doer microrole and another argument, coded as a P argument, expresses an object the doer creates or a situation the

doer is involved in". Despite differences in their grammar, (3.1a-c) are manifestations of this construction type.

(3.1a) *naxud.pelu kārd-en*
pea.cooked.rice do.PST-3PL
'they (the women) made rice with peas' (Ziyarati, Shokri et al. 2013:82)

(3.1b) *tum=ne kya kiyā?*
2SG=ERG what do.PRF.FSG
'what did you do?' (Hindi, Premchand 2017[1936]:5.61)

(3.1c) *guru kicco ker-l-o?*
teacher what do-PST-MSG
'what did the teacher do?' (Kupia, Christmas & Christmas 1973a:75)

Another example for a construction type attested with "do/make" is characterized in Chapter I as uses where "cognate 'do/make' verbs are accompanied by a doer participant, coded as an A argument, an affected participant, coded by an NP flagged by a comitative marker, and a deed element, encoding the type of situation expressed by the clause". This construction type is demonstrated in (3.2a-c) below.

(3.2a) *ab tum=se bahs kaun kar-e bhai!*
now 2SG=COM dispute who do-3SG Brother!
'Now who will argue with you, brother?' (Hindi, Premchand 2017[1936]:3.22)

(3.2b) *bak ləgot judo kor-i bak mar-ise.*
tiger COM fight do-NF tiger die-PST
'he fought with the tiger and killed it' (Nagamese, Sreedhar 1985:194)

(3.2c) *mard-ēw ke abāg zan ī xwēš pašn kun-ēd*
man-INDEF REL with woman LNK REFL contract do.PRS-3SG
'a man who makes a contract with his wife' (Middle Persian, RAF Q62)

Comparative concepts cannot be true or false (as defined by Haspelmath 2010), only more or less useful for comparison. For example, one can choose to not further specify the flagging of the oblique objects in (3.2a-c) as flagged by a comitative marker, and opt for a more abstract characterization of them as oblique objects only. This is a completely acceptable construction type, and would capture the examples in (3.2a-c) and many more.

The issue with generalizing across different types of oblique objects, essentially treating oblique objects in Indo-Iranian as a waste-basket category, is that many Indo-Iranian languages systematically distinguish several oblique objects types. Hindi, for example, distinguishes at least the object type illustrated by (3.2a) above from that in (3.3) below where *gārb-õ* 'poor' is flagged by *par* 'on'. Many construction types in Chapter IV are distinguished by the type of flagging associated with their oblique objects. Thus, lumping all the oblique objects together as a single construction type would create a somewhat inaccurate sense of grammatical uniformity across the family, and would not be profitable in describing the crosslinguistic and intralinguistic variation found in the use of "do/make", "be/become", and "give".

(3.3) *kab^{h}i.kab^{h}i garīb-õ par dayā kiyā*
sometimes poor-MPL on mercy do.PRF.FSG
'sometimes, he took pity on the poor' (Hindi, Premchand 2017[1936]: 5.61)

Apart from their function as the main verbal predicate in the clause, "do/make", "be/become", and "give" often function as auxiliary verbs. Notwithstanding various definitions found in the literature (cf. Heine 1993:3-22, Gildea 1998:33), "auxiliary verbs" are defined here following Payne 1997:84-85 as morphosyntactic verbs which encode grammatical categories such as TAM or other grammatical categories. The construction types in which "do/make", "be/become", and "give" function as auxiliary verbs are differentiated here based on the grammatical category (i.e., defined as a comparative concept) the construction encodes.[25]

The definition of auxiliaries as "morphosyntactic verbs" means that erstwhile auxiliary verbs that underwent reduction and are affixed to the main lexical verb are no longer counted as auxiliary verbs. Consider, for example, the function of "be/become" in expressing the past tense in Middle Persian. To communicate the function of 'past tense', the present tense form of "be/become" is preceded by the main lexical verb. For instance, in (3.4) the auxiliary verb *hēm* 'I am' is found after the main lexical verb *mad* 'came', and the auxiliary construction indexes the S argument. The valence and argument structure of the entire clause are determined by the main lexical verb, and the auxiliary serves to express tense and is the locus for argument indexation.

(3.4) *ka ō ānōh frāz mad h-ēm*
when to there forth come.PST be.PRS-1SG
'when I came there' (Middle Persian, AWN 4.5)

[25] Such a characterization ignores issues like the relative word order of the main and auxiliary verb, the morphological form of the main lexical verb (e.g., finite vs. non-finite forms), and the degree of cohesion between the main and the auxiliary verbs. While Indo-Iranian auxiliary constructions do differ along those lines, adding these parameters in would not be comparatively profitable.

By Early New Persian, the cognates of this Middle Persian auxiliary construction underwent morpho-phonological erosion, and the auxiliary ended up as a suffix in the past tense verb form across all varieties of New Persian, including Early Judeo-Persian (which is a part of the data analyzed in this study). By the stage of New Persian, these suffixes are no longer identifiable as verbs and thus the Early Judeo-Persian Past Tense verb form is not considered an instance of the same construction type as the Middle Persian Past Tense auxiliary construction.

Some auxiliary construction types identified in this dissertation interact with the valence and structure of the entire clause. Consider, for example, the use of "give" in constructions expressing permissive modality, identified in several modern Indo-Iranian languages (see Butt 1995 for a very detailed account of this construction in Urdu). This construction type is demonstrated in (3.5) from Gujarati, where "give" is preceded by the main lexical verb *awwa* 'come'. The valance and argument structure of the clause are jointly determined by the two verbs: "come" donates the microroles of figure, "a broken jug (lit. a jug with a hole)" and ground, "your hand"; while "give" donates the causer, i.e., the person allowing or "letting" the event unfold. As both the auxiliary and the main lexical verb jointly determine the argument structure of the entire clause, Butt 1995 (and many others) analyzed this construction as a type of Verb-Verb complex predication (see Chapter VII for definition of complex predication).

(3.5) *kan-o* *loʈ-o* *pən* *mê* *ta-ra* *hath-mã*
hole-MSG jug-MSG even 1SG.ERG 2SG-GEN.SG hand-LOC
aw-wa *nə* *di-dh-o*
come-INF NEG give-PST-MSG
'I did not let even a jug with a hole come into your hand!' (Gujarati, Christian 1987:242)

All together, over sixty distinct construction types involving "do/make", "be/become", and "give" are identified in this study, and these construction types are presented in Chapters IV through VI. The cross-linguistic frequency of construction types and the number of different languages in which each construction type has been identified interestingly show an almost Zipfian distribution. This is illustrated in Figure 3.1, where the X-axis represents the number of languages in which a construction type is attested, and the Y-axis represents the number of distinct construction types. The X-axis in Figure 3.1 runs from 1, for construction types which are attested in only one language, to 16, for construction types which are attested in all languages in the data. The Y-axis runs from 0 to 20.

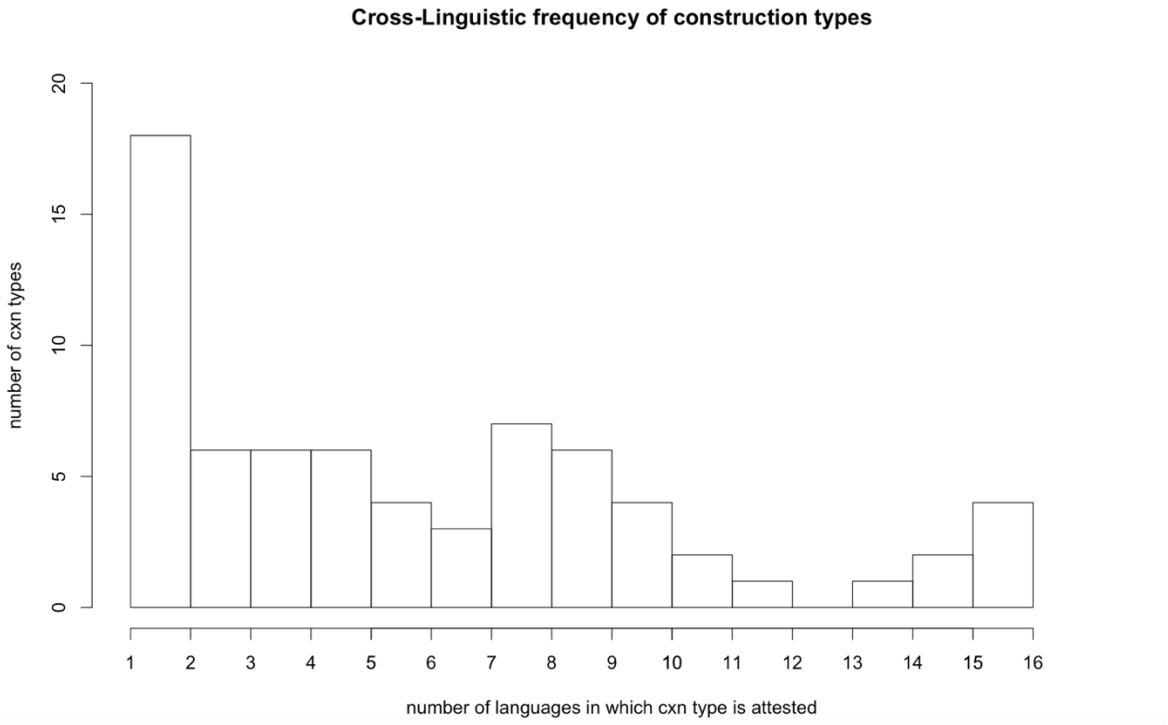


Figure 3.1: Crosslinguistic frequency of construction types for Indo-Iranian

Figure 3.1 shows that a large number of construction types, 17, are attested in only one language and have not been identified in any other language in the data. Six construction types are attested in two to five languages, and seven to three construction types are attested in seven to ten languages. On the right edge of the X-axis, there are four construction types attested in all sixteen languages of the sample and three construction types attested in fifteen languages.

The construction types attested in all sixteen languages are “be/become” construction types used to express core nominal predication, predicate locative, and the existential. The construction types attested in fifteen languages are the common, basic, construction types of “do/make”, such as those illustrated by examples (3.1a-b) above. These construction types are not attested only in Palula, where cognates of Proto-Indo-Iranian “do/make” are not attested.

CHAPTER IV

CONSTRUCTION TYPES OF "DO/MAKE" COGNATE VERBS

4.1 Preliminaries

This chapter describes and analyzes the uses of cognates of PIE **k^{w}er* 'to do, make, construct' (Pokorny 1959:641) in naturalistic texts in the sample of sixteen Indo-Iranian languages used for this study. Cognates of PIE **k^{w}er* are found in fifteen of the sixteen languages studied here. In Palula, an Indo-Aryan language of Northern Pakistan, no cognates of this morpheme were identified.

In some construction types where "do/make" is the main verbal predicate, one could justifiably argue for a Noun-Verb complex predicate (henceforth N-V complex predicate) analysis. Many of these construction types are those in which "do/make" is accompanied by three nominal elements, one of which encodes the type of situation expressed in the clause (see examples throughout this chapter). The complex predicate status of some of these construction types is discussed in Chapter VII, where (following Mohanan 1994, 1997, Haig 2002, Liljegren 2010, *inter alia*) it is argued that N-V complex predicates do not form a unified "family of constructions" and have arisen via several distinct pathways. In this chapter, the analysis is mostly agnostic as to whether some instances of "do/make" construction types are N-V complex predicates or not.

4.2 "Do/Make" as the Main Verbal Predicate

This section analyzes those construction types in which "do/make" functions as the main verbal predicate. These construction types vary in terms of both form and function. On the formal pole, they differ in the number and coding of nominal elements

associated with the verb. On the functional pole, they express a wide variety of situation types, and the participants may be associated with distinct semantic microroles.

Cognate "do/make" verbs are found in combination with one to four nominal elements that encode different components of the situation expressed in the clause. These components may include the propositional contents the clause expresses (state or event) and the different participants. This section is divided into subsections based on the number of elements "do/make" is associated with. Each subsection is further divided according to different construction types. The four main subsections are not equivalent in length, because, perhaps not surprisingly, there are not many construction types in which "do/make" is associated with either one or four elements. The number of construction types in which "do/make" is associated with two elements is slightly higher, and the number of construction types in which "do/make" is associated with three elements is much higher.

4.2.1 "Do/Make" Construction Types with One Nominal Element

There are two construction types in the data in which "do/make" is accompanied by a single nominal element. These construction types are rare both in terms of frequency in texts as well as in terms of the number of languages in which they are found. Their cross linguistic distribution seems limited to Kupia, Kotia Oriya, and Darai, all Eastern Indo-Aryan languages.

4.2.1.1 Temperature Expressing "Do/Make" Clauses

In the first construction type with one nominal element, found only in Kotia Oriya and Kupia (two Eastern Indo-Aryan, Odian, languages), "do/make" functions as the main verbal predicate in clauses expressing temperature. In this construction type, the nominal element expresses some degree of temperature. In all examples in the data, this construction type expresses cold temperature, demonstrated in (4.1) from Kotia Oriya.

(4.1) *pond mas-e bont kakor ko-la*
December month-LOC very cold do-PST.3SG
'It was cold in the month of December' (Kotia Oriya, Gustafsson 1973a:167)

The single obligatory NP in this construction is unflagged, and when the verb exhibits indexing (as in the example above), it indexes the 3rd person singular.[26] As none of the tokens of this construction type contains an expletive element, which would have formally made it a construction type where "do/make" is associated with two nominal elements, the verb can be argued to be indexing the nominal element expressing the temperature, formally a 3rd person singular form. This usage is similar in some respects to Spanish *hace frío* "it's cold", literally "(it) does cold".[27]

[26] The frequency of main verbal predicates which do not index an argument in Kotia Oriya and Kupia seems very high when compared to other Indo-Iranian languages.
[27] I thank Manuel Otero for pointing this out to me.

4.2.1.2 "In Particular" Construction

The second "do/make" construction type with a single nominal element has a non-finite form of "do/make" (analyzed in Dhakal 2013 as the "Sequential" form), preceded by *kʰas* 'in particular, especially' and a NP. The NP is the Topic (and Subject) of the inflected verb which immediately follows the non-finite form of "do/make". This construction type is rare in the data, and has been identified only in Darai, an Eastern Indo-Aryan language of Nepal. This construction type is shown in (4.2), from a text describing the traditional preparation method of a beverage, for which the leaf of a specific tree is required (but as the following clause lets us know, in its absence the leaf of another tree will also do):

(4.2) *pat kʰas kər-ikun bʰãḍi-kə cah-i-t*
leaf especially do-SEQ bhandi.tree-GEN need-PRS-3SG
'the leaf is required especially of the bhandi tree' (Darai, Dhakal 2013:63)

Kʰas 'in particular, especially' seems to highlight the information status of the preceding NP, in (4.2) *pat,* 'leaf'. This construction type is always translated in Dhakal 2013 by using the adverbs "especially" or "in particular". Its function, then, can be tentatively argued to signal the importance or centrality of the sole argument of "do/make" (as opposed to both other parts of the tree as well as leaves of other trees) in the process described by that specific stretch of discourse. As this construction is rare in the data, it is impossible to give a more elaborate and convincing analysis of its function and clausal status.

4.2.2 Construction Types with Two Nominal Elements

Construction types where "do/make" is accompanied by two nominal elements are both more common and more varied than those in which it is accompanied by one element. First, they are more common in terms of the number of languages in which they are attested. While construction types where "do/make" is accompanied by one nominal element are attested only in three (out of sixteen) languages, construction types where "do/make" is accompanied by two nominal elements were identified in the all fifteen languages where cognates of "do/make" were actually retained. Second, construction types where "do/make" is accompanied by two nominal elements also have higher token and type frequency.

The construction types presented in this subsection differ in the coding of the elements associated with "do/make". This includes morphological flagging, indexation, and relative word order. In some construction types, "do/make" expresses the primary lexical semantics (i.e., events of creation). In others, the lexical semantics of the predication are encoded by the semantics of one of the nominal elements associated with "do/make." Whether these are instances of N-V complex predication or not depends on how one operationalizes the definition of complex predication; see Chapter VII for details. In yet other construction types, the lexical semantics of the clause is encoded holistically by the clause constructions (e.g., "do/make" as sole verbal predicate in motion expressing clauses). Finally, the construction types presented in this subsection differ in their frequency. Only two construction types are found in all fifteen languages where cognate "do/make" verbs have survived. The other construction types are only found in one to three languages in the sample.

4.2.2.1 "Do/Make" Accompanied by A and P Arguments

In this construction type, "do/make" is accompanied by two nominal elements. In this subsection I refer to these elements by the microroles "doer" and "deed", as they encode the doer of the action and the deed accomplished, or the object they create, respectively. In this construction type, doers are coded as A arguments and deeds / objects created are coded as P arguments, and the clause expresses the creation of the object (P argument) by the doer (A argument), or that the referent expressed by the A argument is performing some action expressed by the P argument.[28]

In examples (4.3a-b) the doer participant is expressed as an A argument. This is evident, for example, by the indexing of the doer participant on the verb. The P argument expresses an object which was prepared or made by the referent of the A argument. In (4.3a) it is a dish of cooked rice and peas, and in (4.3b) it is a wooden horse. It should be noted that in most of the modern Indo-Aryan languages in the sample, there is at least one competing verb with similar semantics, such as Hindi verb *banā* 'make', which often expresses events of creation such as the two in (4.3a-b). Such verbs seem to be slowly replacing the *kar* type "do/make" cognate verb in expressing events with creation or manufacturing semantics.

(4.3a) *naxud-pelu* *kārd-en*
pea-cooked.rice do.PST-3PL
'they (the women) made rice with peas' (Ziyarati, Shokri et al. 2013:82)

[28] There are some differences in behavioral properties between deed elements which express an event or situation and deed events that express an object created by the doer.

(4.3b) *goʈek boɽoi goʈek goc-or goɽa goʈek ko-la*
INDEF carpenter INDEF wood-GEN horse INDEF do-PST.3SG
'One carpenter made a wooden horse' (Kotia Oriya, Gustafsson 1973a:190)

In examples (4.4a-b) below, the doer and the deed elements are also expressed as A and P arguments respectively. The P argument in these examples expresses the type of action or event performed by the A argument, rather then a type of object the A argument creates. The events expressed by clauses similar to those in (4.4a-b) below are often semantically intransitive events: apart from the NP expressing the action accomplished there is only one more associated argument.

The semantically intransitive nature of these events, however, does not lead to an intransitive pattern in argument structure. In example (4.4a), from Ziyarat Mazandarani, the deed element is coded like a P argument, and is even flagged by the Ziyarati Differential Object marker =*re*. In Ziyarat Mazandarani, DOM markers are associated with identifiable, specific P referents. Thus, they are rarely found flagging P arguments that express types of events for obvious semantic reasons. In (4.4a), however, the discussion revolves around who does certain jobs in certain times, and hence the type of situation expressed by there is more specific and identifiable.

Similarly, example (4.4b), from Hindi, expresses a semantically intransitive event ("protest"), but the argument structure is identical to the one used to encode prototypical transitive events. The doer, *Gobar* is coded like an A argument, and is flagged by the Hindi Ergative marker =*ne*. Similarly, the deed *prativād* 'protest', a feminine singular noun, is coded like a P argument and is indexed by the verb in number and gender.

(4.4a) *ham dāmdāri=re hā-kān-en ham kešāvarzi=re*
also cattle.breeding=OBJ PRV-do.PRS-3PL also agricultural.work=OBJ
hā-kān-en
PRV-do.PRS-3PL
'(the women of Ziyarat) … they also breed cattle and also work in the fields…'
(Ziyarati, Shokri et al. 2013:206)

(4.4b) *gobar=ne prativād kiyā*
gobar=ERG protest do.PST.FSG
'Gobar protested' (Hindi, Premchand 2017[1936]: 3.22)

The examples given in this section so far demonstrate the two semantic functions of this construction type: in the first, P arguments express tangible objects created by the referent of the A argument and in the second they express some sort of action or event (or rarely, state) performed by the referent of the A argument. These two types differ in tendencies of P-argument coding, but these are not systemic grammatical differences, which would have necessitated differentiating two construction types.

One difference between these two semantic types of deed elements is that deed elements encoding events and states are less frequently expressed by semantically "light" NPs such as demonstratives and question words. This, however, is simply a tendency, as examples (4.5a-b) below demonstrate. In both examples, the P argument expresses some sort of an event. The speaker in example (4.5a) asks what should he and his companions do in the context of their current situation. In (4.5b), the subject of the clause will perform something, mentioned as a contextually retrievable zero, in secret (literally "while hiding").

(4.5a) *vāt=eš hālā če be-ker-ime?*
say.PST=3SG now what IRR-do.PRS-1PL
'he said: "what shall we do now?"' (Sivandi, Lecoq 1979:90)

(4.5b) *vah cchipa-kar kar-e-g-ī*
DEM hide-CV do-3SG-FUT-FSG
'she will do it hiding' (Hindi, Premchand 2017[1936]: 5.63)

Another possible difference between these two types concerns the coding properties of deed elements, particularly whether DOM is used. Across Indo-Iranian, the flagging of the P argument is often correlated with more identifiable and specific P arguments. When the deed element expresses an event or situation, it is less likely to be identifiable or specific, as it expresses a type of an event or state, which are typically non-referential and hence not identifiable or specific. Thus, deed elements expressing situations are less likely to be flagged by an object marker. As seen in (4.4a) above, however, the expression of events as P arguments in this construction type is sometimes (albeit very infrequently) flagged by an object marker.

The low frequency of DOM and the rare use of semantically "light" NPs (such as "what" or demonstratives) for events expressed as P arguments in this construction type are, however, tendencies and not categorical patterns. Whether these different tendencies justify splitting this construction type into two is a matter of descriptive preferences. If one decides to split this construction type based on the semantics of the deed element (i.e., whether it expresses an object created or an event or situation), both construction types would be attested in the fifteen languages in which cognate "do/make" verbs are attested.

4.2.2.2 Experiencer and S/A Arguments

In this "do/make" construction type, it is found together with two nominal elements. The first of these expresses a semantic experiencer participant, and is not coded like A/S arguments. The other argument expresses the state experienced and is expressed by an unflagged NP, which status I analyze here as a (less prototypical) S/A argument. This construction type is found only in Nagamese and is demonstrated in (4.6). In this example, the experiencer argument is flagged by the Nagamese Differential Object marker =*ke*, which is used to flag identifiable primary objects (P and R arguments). The state experienced is coded by the morphologically unflagged NP, *bhorta* 'satisfaction'.

(4.6) *apuni-khan=ke bhorta kor-ibo!*
2-PL=OBJ satisfaction do-FUT
'you will be satisfied (Lit. Satisfaction will do you)' (Nagamese, Luke 6:21)

The status of *bhorta* 'satisfaction' in example (4.6) is somewhat difficult to pin down with confidence. First, as Nagamese does not exhibit argument indexation on the verb, the only relevant coding properties are flagging of arguments and relative word order. Nagamese S/A arguments are unflagged, while P/R arguments are often flagged by the DOM =*ke*.[29] When it comes to word order, however, the usual word order is A-P-verb.

[29] Nagamese =*ke* flags R arguments in general, with no sensitivity to identifiability. This is similar to many primary object markers across Indo-Iranian, which are used to flag P arguments under DOM but R arguments more generally.

This means that examples like (4.6) have an unusual word order as far as Nagamese word order tendencies are concerned, which can tentatively be explained by alluding to the non-prototypical semantics of the S/A and P/R arguments in this construction. For the most part, S/A arguments tend to be animate, referential, and established discourse entities, whereas P/R arguments tend to be inanimate, and less established discourse entities. In this construction type, however, it is the P/R arguments which are animate and established in the discourse, and the S/A arguments are inanimate, non-referential, and not established in the discourse. In terms of continuity in the discourse, the P/R argument in (4.6) above is a discourse participant, and is referred to both before and after 4.6. In contrast, the S/A arguments in (4.6) is not referred to again in discourse. Thus, it is at least a plausible hypothesis that the unusual word order in this construction reflects the relative discourse prominence of the two arguments, and not their syntactic status.

4.2.2.3 S Argument and a Locative Element

This third "do/make" construction type with two nominal elements is rare, and is found only in Hindi. The first NP is coded as an S argument,[30] while the second NP is flagged by the Hindi Locative marker *=mẽ*. Despite the locative flagging, the clause does not express motion or location, but instead indicates an unwanted quality of an event. In (4.7), the speaker, i.e., the boy's father, says that there must be no delay in the engagement of his child. The S argument is *der* 'delay, lateness' and the locative

[30] At least, there is no positive evidence that this participant is ever coded as an A argument (e.g., flagged with the Ergative marker *=ne*).

argument is *laɽke ki sagaːiː mẽ* ‘the engagement of the boy’. The verb seems to be indexing *der* ‘delay’, a feminine singular NP.[31]

(4.7) *ab laɽke ki sagaːiː mẽ der na kar-n-i caːhiːye*
now boy.OBL GEN engagement LOC delay NEG do-INF-FSG must.PRS
‘the boy’s engagement should be soon’ (Hindi, Premchand 2017[1936]: 4.34)

4.2.2.4 A/S Argument and an Oblique Comitative Object

In this “do/make” construction type with two elements, the first element is coded as an S/A argument and the second is a NP flagged by the comitative marker, i.e., an oblique object. This construction type is rare, and is found in the data only in Gorani. The NP flagged by the comitative marker expresses the way in which the S/A argument behaves or needs to behave (when the verb form indicates modality). In (4.8), the S/A argument is the second person pronoun *tu,* and the manner in which its referent needs to behave is expressed by the PP *wa qawłim* ‘with my words’.

(4.8) *eh tu wa qawł=im bi-ka*
ugh 2SG with word=1SG SBJ-do.IMPR
‘ugh, do what I say!’ (Gorani, Mahmoudveysi et al. 2012, 5:62)

The exact function or the functional difference between this construction type and the one in which “do/make” is accompanied by A and P arguments is unclear. As

[31] This construction type was only found in the data a handful of times, all in a non-indicative modal context. However, a web search of modern Hindi data found further instances of this construction type, the first twenty of those in a non-indicative context as well.

example (4.9) below shows, the same semantic content, “my words,” can be found without flagging by the preposition *wa* ‘with’, as well.

(4.9) *ara qawł=im bi-kar-ī*
if word=1SG SBJ-do.PRS-2SG
‘(you can,) if you do what I say’ (Gorani, Mahmoudveysi et al. 2012, 4:198)

4.2.2.5 Copula(-like) Usages

This fifth “do/make” construction type with two nominal elements is rare and has been identified only in Early Judeo-Persian. This construction type expresses equation on some dimension. In (4.10) below, the clause expresses the translation equivalence of the two arguments, expressing words in Biblical Hebrew and Early Judeo-Persian. This construction type is similar in form and function to the English *two and three make(s) five.* Example (4.10) predicates the equivalence between the Hebrew word written *ʕl* (Biblical Hebrew *ʕal,* Modern Hebrew *ʔal)* and the Early Judeo-Persian word written *ʔbr* (Middle Persian *ābar*). Both words are prepositions, meaning, roughly, ‘on’.

(4.10) *ʕl ʔbr kwnd*
on(Heb.) on do.PRS.3SG
‘(The Hebrew word) ‘ʕl’ means “on”’ (EJP, Gindin 2008, 1:10)

The relative rarity of this construction type in the data can be a result of the distinct genre of the Early Judeo-Persian texts (i.e., exegesis of Ezekiel) compared to the genres of the texts used in other languages (i.e., mostly short narratives, interviews, and descriptions of cultural events). However, at least parts of the Middle Persian and

Sogdian data used for this dissertation are similar in genre to the data from Early Judeo-Persian, and in the data from these other languages, this construction type is missing.

4.2.2.6 "Do/Make" in Motion Constructions

This subsection describes a "do/make" construction type with two elements in which the clause expresses motion. "Do/make" is accompanied in this case by two arguments expressing the moving figure and the goal of motion. As none of the elements accompanying "do/make" in this construction type can be considered encoding the type of situation expressed by the clause, this construction type cannot be considered to be an instance of N-V complex predication.

This construction type is illustrated in (4.11a-b). In (4.11a), from Gorani, the moving figure is expressed by *gala* 'flock (of sheep)' and the goal of motion is expressed by the post-verbal NP *ā das* 'that side (of the river)'. The verb is followed by the directional verbal clitic *=ya,* one of the unique features of Gorani in the data sample here,[32] which is often found following the verb in clauses expressing motion, caused motion, or transfer. The post-verbal position of the goal of motion is its common position in many Central Iranian languages, which are otherwise mostly verb-final. Thus, apart from the verbal lexeme, the clause grammar of examples like (4.11a) is congruent with the grammar of motion events that are expressed by lexical motion verbs. In (4.11b), from Nagamese, the moving figure is expressed by the unflagged NP *john* and the ground is coded by the post-positional phrase *Jordan nodi par* 'on/at the Jordan river'.

[32] The exact origin of the directional marker is somewhat unclear (Mahmoudveysi et al. 2012:50-51).

(4.11a) *gala ma-kar-ī=ya ā das*
flock IND-do.PRS-3SG=DRCT DEM side
'the flock went to the other bank' (Gorani, Mahmoudveysi et al. 2012, 2:7)

(4.11b) *itu nimite john jordan nodi par kor-ikene*
DEM for John Jordan river on do-TEMP
'then, while John was going (from village to village) on the Jordan river'
(Nagamese, Luke 3:3)

Clauses expressing motion, or change of location, without a lexical motion verb or any verb at all, are not unique to Gorani or Nagamese. Example (4.12), from Middle Persian, comes from a portion of a text describing the process that souls undergo after death. It expresses the destination of souls of those people who are said to have much merit, as opposed to those who were sinners. The destination of the pious is expressed by the prepositional phrase *ō wahišt* 'to heaven'. Such examples cannot be argued away as instances of verb ellipsis: in the immediate, and the not-so-immediate, vicinity of these clauses there is no appropriate motion verb, or, indeed, any other appropriate verb.

(4.12) *harw ke=š sē srōš-čarnām kirbag wēš ku wināh ō wahišt*
every REL=3SG three srosh-charanam merit more COMP sin to heaven
'everyone whose meritious deeds are three srosh-charanam greater then his sins, will go to heaven' (Middle Persian, AWN 6.5)

Such examples, in tandem with the Gorani and Nagamese examples presented here, suggest that in at least some Indo-Iranian languages, motion semantics can also be expressed by clause level constructions without the lexical semantics associated with the

use of a motion verb. That is, motion is not encoded by any single constituent or component of the clauses in examples (4.11 – 4.12), but only by the ensemble of components in the clause.

4.2.2.7 "Do/Make" in Speech Constructions

In one construction type, found in Kupia and Kotia Oriya, "do/make" is found in clauses expressing speech events which include direct or indirect quotes. The function of "do/make" in these clauses can be interpreted in two ways. One interpretation is that it is the only verbal predicate in clauses expressing a speech event, and the other is it has (sometimes) grammaticalized together with a speech verb as a part of a quotative marker. In either analysis, this function of "do/make" forms a unique construction type, and is a result of an interaction between two grammaticalization processes: speech verbs grammaticalize to function as quotation markers and "do/make" grammaticalizes as an auxiliary verb (see below for an analysis of "do/make" in its auxiliary function). In examples (4.13a-b), "do/make" is preceded by direct quotes, ending with a quotation marker. Given the glosses, in both examples "do/make" seems to be the only verbal predicate in clauses expressing direct speech.

(4.13a) *a:ŋwu taruwata je-yinde mena ker-a*
1SG then come-FUT.1SG QUOT do-NF
'I said "then, I'll come"' (Kupia, Christmas & Christmas 1973a:41)

(4.13b) *a=re baia suntia tui posu za boli kor-i*
oh=2SG crazy shrew 2SG mindless.animal go.IMPR QUOT do-NF
'he said "oh you crazy shrew, you mindless animal! Go away!"' (Kotia Oriya, Gustafsson 1973a:171)

This interpretation of (4.13a-b) is not, however, unproblematic. The origin of both quotation markers, *mena* in Kupia and *boli* in Kotia Oriya, is the non-final form of "say". A clear illustration of the usage of *mena* and *boli* in the function of a quotation marker, when they are accompanied by another speech verbs is in examples (4.14-4.15) below. In examples (4.14a-b), *boli* and *mena* are used following a direct quote which is not a part of a speech act at all. In examples (4.15a-b), *boli* and *mena* are used as quotation markers in clauses expressing speech events. In these clauses, *boli* and *mena* are followed by the A argument of the clause referencing the speaker, and the main verb of the clause, which is a verb of speaking.[33]

(4.14a) *birat des-e podmusen boli goṭek raza roi-la*
birat land-LOC podmusen QUOT INDEF king remain-PST.3SG
'In the country of Birat, there was a king named Podmusen' (Kotia Oriya, Gustafsson 1973a:111)

(4.14b) *tollito jalapuṭu ayitar santa mena etki ji:ni ku:ɖanu gec-umde*
first Jalaput Sunday market QUOT all people indeed go-FUT.1PL
'first, speaking about the Jalaput Sunday market, we all go' (Kupia, Christmas & Christmas 1973a:17)

[33] The grammaticalization of speech verbs as quotative markers is a common grammaticalization pathway, cross-linguistically well attested, and is also found in other Eastern Indo-Aryan languages such as Bengali.

(4.15a) *a' de-bi boli raza koi-la*
yes give-FUT.1SG QUOT king say-PST.3SG
'"yes, I will give it" said the king' (Kotia Oriya, Gustafsson 1973a:191)

(4.15b) *jo=wi am=ka ka:wale mena tu:mu saŋ-a nay*
DEM=EMPH 1PL=OBJ required QUOT 2SG say-NF NEG
'don't say "he's the one we want"' (Kupia, Christmas & Christmas 1973a:60)

The problem with the interpretation of examples like those in (4.13a-b) above lies in the fact that, while there is no doubt that *mena* and *boli* have grammaticalized into quotation markers, shown by (4.14 –4.15), these forms can also function as the only verbal predicate in clauses expressing direct and indirect speech, as in (4.16a-b). Such examples are relatively rare in both languages, but are more common in Kotia Oriya.

(4.16a) *cappuna aŋ=ka wa:li mull-a de:-su men-a*
completely 1SG=OBJ string leave-NF give-IMPR say-NF
'"leave me the whole string", she said (and the queen had done so)' (Kupia, Christmas & Christmas 1973a:90)

(4.16b) *u' ru-a bol-i goʈek gor de-lai*
yes remain-IMPR.PL say-NF INDEF house give-PST.3PL
'"yes, stay!" he said, and they gave us a house' (Kotia Oriya, Gustafsson 1973a:115)

Thus, the forms *mena* and *boli* have (at least) two functions:[34] as speech verbs and as grammaticalized quotation markers. In examples like (4.15a-b), where *mena* and *boli* are followed by a NP representing the speaker and a speech verb, their function should be interpreted as quotation markers. In examples like (4.16a-b), however, the function of *mena* and *boli* should be interpreted as a verb of speaking, as there is no other verbal predicate in the clause.

Returning to (4.13a-b), it could be suggested that the sequences *mena kora* and *boli kori* are instances of the "do/make" auxiliary construction type, with a speech verb as the main lexical verb. While this interpretation is possible in at least some instances of *mena kora* in Kupia and *boli kori* in Kotia Oriya, there are two types of situations which raise doubts about it.

In (4.17a-b) "do/make" follows a speech verb, and this sequence can be interpreted as an instance of the "do/make" auxiliary construction. In both examples, the speech verb and "do/make" are followed by an overt mention of the speaker and a lexical verb of speech. Thus, examples (4.17a-b) below are parallel to (4.15a-b) above, where *boli* and *mena* were followed by a mention of the speaker and a lexical speech verb. The difference between the two pairs of examples is that in (4.17a-b) below, "do/make" seems to form a part of the quotation marker together with *boli* and *mena*.

[34] The retention of the "heavy" function of being a speech verb together with the grammaticalized function of quotation marker is not rare, and is found also in Sinitic (Chappell 2008).

(4.17a) *sedi sobu lok=ke boḍ lok bol-i kor-i pani koi-la*
DEM all person=OBJ big person say-NF do-NF water say-PST.3SG
‘“the rain cloud is a great person to all people” said the water’ (Kotia Oriya, Gustafsson 1973a:170)

(4.17b) *ja kaɳtapuri paṭṭeḍḍa aŋ=ka ka:wale men-a ker-a*
DEM kantapuri necklace 1SG=OBJ required say-NF do-NF
i:njo guru saŋ-ede
DEM teacher say-FUT.3SG
‘The guru will say to her “I want the Kantapuri necklace”’ (Kupia, Christmas & Christmas 1973a:88)

The second problem with an interpretation of “do/make” as an auxiliary emerges in clauses in which there is a mismatch between the function of the “do/make” auxiliary construction (as sketched in later in this chapter) and the use of that specific clause. Roughly, the function of “do/make” auxiliary constructions in Kupia and Kotia Oriya is to signal that two events are thematically related, but distinct. In the main event line of narratives, these constructions express a non-simultaneous relationship between two events (although they can express simultaneous states outside of the main event line). Some instances of *mena kera* in Kupia and *boli kori* in Kotia Oriya in the main event line of narratives cannot be interpreted as expressing sequential events.

In example (4.18) below, from Kupia, there are two clauses and a direct citation. Clause (a) expresses a speech event and the verbal predicate in this clause is *mena kera*. Clause (b) expresses the function of the speech event in the conversation it is a part of, and the main verbal predicate is a finite form of “do/make”. In this example, an evil teacher is trying to deceive the king by telling him that the king’s older son learned how

to read, whereas it was the younger son who was taught to read, while the teacher assigned the older one housework. The event of deception and the speech event are the same event. That is, *mena kera* does not code a distinct event from the one in the following clause, as would be expected were it the "do/make" auxiliary construction.

(4.18) (a) *well-o=wi ba:ga sadu ker-a as-e men-a ker-a*
big-M=FOC thoroughly reading do-NV be-PRS.3SG say-NF do-NF
(b) *i:njo ra:n-o=ka mo:simi ker-l-o*
DEM king-M=OBJ deception do-PST-3MSG
'(a) he said: "the big one is reading very well!" (b) thus deceiving the king'
(Kupia, Christmas & Christmas 1973a:62)

As shown above, there is some quirky interaction between the grammaticalization of *mena* and *boli* as quotation markers and the auxiliary construction with "do/make" auxiliary. This interaction results in one of two distinct and unique construction types: (a) "do/make" is used as the only verbal predicate in clauses containing a speech event, or (b) the "do/make" auxiliary construction has grammaticalized as a part of a quotation marker. It is difficult to distinguish these two options from one another, but whichever analysis is more appropriate, this construction type is distinct from the others identified in elsewhere in this study.

4.2.2.8 Short Summary

Subsection 4.2.2 has described the different construction types in which "do/make" appears together with two nominal elements. There are seven such construction types in the data which always differ from each either in form and function.

In some of these construction types, the lexical semantics expressed by the clause is encoded by either "do/make" or by a nominal element in the clause. In other construction types, the semantic meaning expressed by the clause is not encoded by any specific constituent, but by the clause construction as a whole. Finally, the analysis of at least one construction type discussed above, i.e., in which "do/make" is accompanied by a non-finite form of "say", is uncertain and it could very well be that "do/make" functions there as a part of a (complex) quotation marker.

4.2.3 Construction Types with Three Elements

This subsection describes construction types in which "do/make" is accompanied by three elements. These elements are referred to here by the microroles of doer and affected participants, and the deed element. The doer is the participant performing the action or controlling the situation expressed in the clause. The deed element refers to the nominal, or very rarely adpositional, element encoding the type of situation expressed in the clause. The affected participant refers to the participant affected by the event or situation expressed in the clause. The doer, deed, and affected microroles do not always correspond to participants with a syntactically or semantically unified semantic role or grammatical relation, and their coding and semantic properties vary across construction types.

Many instances of these construction types in the languages studied here have been analyzed as N-V complex predication (Mohanan 1994, 1997, Haig 2002, Samvelian 2012, *inter alia*). One of the main problems associated with the analysis of complex predication is the syntactic status of the element which encodes the type of situation the

clause expresses, here referred to as the deed element. Throughout this section, some of the problems raised by the analysis of these elements will pop up again and again. There have been many debates in the literature about their analysis and their relationship to the verb, especially in New Persian (e.g., Goldberg 2003, Müller 2010, Samvelian 2012), but analyses have also been proposed for at least Kurmanji Kurdish (e.g., Haig 2002), Hindi (e.g., Mohanan 1997), and Palula (Liljegren 2010). The syntactic status of the deed elements is further analyzed in Chapter VII, where it will be argued, following Mohanan 1994, 1997, Haig 2002, and Liljegren 2010, that N-V complex predication in Indo-Iranian does not form a unified grammatical set of construction types diachronically or synchronically. This subsection, however, remains non-committal as to whether a construction type (or an instance of a construction type) could be analyzed as a N-V complex predicate. Hence, some problems of analysis, including the analysis of the status of deed elements, will be mentioned only briefly in this chapter.

4.2.3.1 Doer as A Argument, Affected Participant as P, Deed as NP

In this first "do/make" construction type with three elements, "do/make" is found in clauses expressing many different types of situations: from prototypical transitive events to events of perception, cognition, and utterance (PCU events). "Do/make" in this construction type is accompanied by three nominal elements, expressing the doer participant (coded as an A argument), deed or the event type expressed by the clause (coded as an unflagged NP), and the affected participant (coded as a P argument). This is the most common construction type with three nominal elements in the data, and it is

found in all languages of the sample which retain cognates of "do/make". It will be referred to as the "main three elements "do/make" construction type".

Two initial examples for this construction type are in (4.19a-b), both expressing events of destruction. In (4.19a-b) the deed elements are *barabād* and *tabāh*, respectively, both meaning 'ruin, destruction'. The doer participant is expressed as an A argument, as indicated by the verbal indexing in (4.19a) and by the clitic pronoun in (4.19b).

(4.19a) *hōrī ghar barabād kar dē-g-ā*
Hori house ruin do give-FUT-MSG
'Hori will ruin (his) house' (Hindi, Premchand 2017[1936]: 4.33)

(4.19b) *u=š kodak tabāh kard*
and=3SG child ruin do.PST.3SG
'she destroyed her child' (Middle Persian, AWN 64.5)

Another common semantic domain expressed by this construction type is caused change of state of the affected participant by the doer participant. Here, the deed element expresses the new state, and often denotes a semantic property, usually through a lexical adjective, but also often through a lexical noun. This is illustrated in (4.20a-b) from Hindi and Middle Persian. In (4.20a) the new state is denoted by the lexical adjective (and Persian loan word) *kʰuʃ* 'happy', and in (4.20b) by *garm* 'hot'. The grammatical coding of the doer and affected participants, as well as that of the deed element, is identical to that in examples (4.19a-b) above.

(4.20a) *amalõ=ko* *kʰuʃ* *kar-o*
workers=OBJ happy do.PRS-3SG
'he made the workers happy' (Hindi, Premchand 2017[1936]: 3.22)

(4.20b) *ceōn āhangar* *āhan* *tāb-ēd* *ud* *garm* *kun-ēd*
as blacksmith iron inflame.PRS-3SG and hot do.PST-3SG
'as a blacksmith inflames the iron and makes it hot' (Middle Persian, DK6 E22.f)

Examples (4.19-4.20) above clearly show that the doer participant is coded as an A argument. In (4.19a), the verb indexes the doer, expressed by the proper name *hori*, and in (4.19b) the A argument is expressed by the 3SG Clitic Pronoun. The Middle Persian Clitic Pronoun expresses only A arguments, not P or R arguments, when the main verbal predicate is in the past tense.

Not all of the examples, however, provide clear evidence that affected participants are coded as P arguments. In examples (4.19a-b) and (4.20b), for instance, the context does not allow for a clear and unambiguous identification of which constituent should be considered the P argument: the deed element or the affected participant. In (4.19a), for example, *ghar* 'house' is not flagged by the Hindi DOM marker, and as the clause is not an instance of the Hindi Ergative construction, *ghar* 'house' is also not indexed on the verb. In (4.19b), a Middle Persian Ergative construction, the verb indexes a 3SG element that could be either *kodak* 'child' or *tabāh* 'ruin'. In (4.20b), the P argument (the iron) is not overtly expressed in the clause at all, and the A argument (the doer participant) is indexed on the verb.

DOM and verbal indexation do distinguish, however, the affected and deed elements. One such example is (4.20a) above, from Hindi, where the affected participant

is flagged by the DOM *=ko*. Examples (4.21a-b), from Sivandi and Kotia Oriya, further show the status of affected participants as P arguments. The affected participants, *monajjemga* 'astrologers' in (4.21a) and *se des* 'this land' in (4.21b), are flagged by DOM as specific, highly identifiable P arguments. The doer is not overtly expressed by a NP in either example, but is indexed by the clitic argument pronoun in (4.21a), and on the verb in (4.21b). The deed element is expressed by a morphologically unflagged NP in both (4.21a-b), *ɣond* 'collection' in (4.21a) and *palon* 'inspection' in (4.21b).

(4.21a) *monajjemga=rā ɣond=eš kerd*
astrologer.PL=OBJ collect=3SG do.PST
'he assembled the astrologers' (Sivandi, Lecoq 1979:137)

(4.21b) *se des=ke palon ko-lai*
DEM land=OBJ inspection do-PST.3PL
'they inspected that land' (Kotia Oriya, Gustafsson 1973a:201-202)

The question remains regarding the syntactic status of the deed element. There is no clear evidence to support an analysis of deed elements in this construction type as P arguments, as they are never flagged by object markers and are never indexed on the verb in ergative constructions. Whether the deed elements denote a type of situation as in (4.19, 4.21) or a property concept as in (4.20a-b), their coding properties are identical across instances of this construction type. For now, the syntactic status of these elements is left open, but it should be noted that the similarity in coding properties across instances of this construction type point to the initial stages in one of the diachronic pathways leading to the rise of N-V complex predicate constructions across Indo-Iranian. This is

taken up again in Chapter VII.

The range of semantic situation types in this construction type has been shown in examples (4.19, 4.21). In (4.19a-b), the clauses express highly transitive events, where the doer participant affects the affected participant physically. In (4.21b), on the other hand, the P argument, *se des* 'this land', is not physically affected by the event. In examples (4.22a-b), this construction type expresses events very low on the transitivity scale.

(4.22a) *tin-ʈa am lob kər-lə toi*
three-CLF mango greed do-PST 2SG
'(because) you showed greed for those three mangoes' (Darai, Dhakal 2013:84)

(4.22b) *in kār=(r)e šeru ā-kerd-em,*
DEM deed=OBJ start PRV-do.PST-1SG
'I began this job' (Ziyarat Mazandarani, Shokri et al 2013:205)

4.2.3.2 A Argument, Deed Element, and an (Adnominal) Modifier

In this construction type, "do/make" is used in clauses which express several different types of situations, and is associated with three nominal elements, again referred to here as the doer, affected participants, and the deed element. In this construction type, doer participants are coded as A arguments, and deed elements are morphologically unflagged NPs, which are adnominally modified by a NP expressing the affected participant.

This construction type is less frequent than the main construction type with three elements described in the previous section in three senses. First, it is found in Nagamese,

Hindi, Sanskrit, Pali, Sivandi, and Early Judeo-Persian, whereas the main three-elements-construction type is found in all fifteen languages with cognate "do/make" verbs. Second, in languages in which this construction type is found, it is less frequent than the main three-elements-construction type. Third, it expresses a more limited range of event types. The second and third points might be closely related: the range of event types expressed in this construction type could be narrower simply because the construction type is less frequent, or the construction type could be less frequent because it is associated with a more limited range of event types.

Examples of this construction type are given in (4.23a-b), from Nagamese and Hindi, two Indo-Aryan languages. In both examples the deed element is preceded by an adnominal modification marker (*ki* in Hindi and *laga* in Nagamese), which is itself preceded by the NP expressing the affected participant. In Sanskrit and Pali where this construction type is also found, the adnominal modifier is in the Genitive case.

(4.23a) *jʰigur do hal ki kʰeti kart-aa hai*
Jhigur two plough GEN farming do.PTCP-MSG be.PRS.3SG
'Jhigur farmed two fields' (Hindi, Premchand 2017[1936]: 4.45)

(4.23b) *manu-khan laga phaisla na-kor-ibi,*
person-PL GEN judgment NEG-do-IMPR
'don't judge people' (Nagamese, Luke 6:37)

The coding of the doer participant as the A argument is clear in (4.23a-b). In (4.23a) the doer participant, *Jʰigur,* a proper name, is expressed by an unflagged NP and is indexed on the verb. In (4.23b), the doer participant is the addressee of the imperative

clause. The coding of the affected participant as an adnominal modifier is clear from its flagging by the adnominal modifier markers *ki* or *laga* (both glossed as GEN), in (4.23a-b).

The status of the deed element is less clear in (4.23a-b). In both clauses, it is the only other constituent NP in the clause apart from the doer participant, and could be argued to be the P argument "by default". Clear evidence for the syntactic status of deed elements as P arguments, however, would be found only in clauses where the deed element is flagged by DOM, or in ergative clauses, where it would be indexed on the verb. In no example of this construction type in the data, from any language, was the deed element flagged by DOM. This is not surprising as the deed element never expresses a specific, identifiable entity, which is the type of entities usually flagged by DOM.

The only positive evidence for the status of these deed elements as non-prototypical P arguments is available in languages where verb indexing follows an ergative alignment pattern, at least in some constructions. In (4.23c), a Hindi Ergative construction, the doer participant is flagged by the Hindi Ergative marker *=ne.* The affected participant, *mehma:nõ* 'guests' is expressed as an adnominal modifier of the deed element *svagat* 'greetings'. The verb indexes a 3FSG argument, which in (4.23c) would index only the deed element "greetings", as Roy Sahib is masculine and the guests are plural.

(4.23c) *roy sahib=ne mehma:nõ ka: svagat kiya:*
roy sahib=ERG guest.PL GEN greetings do.PRF.3FSG
'Roy Sahib welcomed the guests' (Hindi, Premchand 2017[1936]: 6.65)

In languages without ergative constructions, there are no similar options to positively indicate the status of the deed element as a P argument. As deed elements in this construction types are also never found flagged by DOM, there is no positive evidence from argument coding for the status of deed elements as P arguments. Thus, there is at least some crosslinguistic variation in whether overt coding evidence is available for the syntactic status of the deed element as a P argument in this construction type. Finally, it should be kept in mind that the deed element in this construction type cannot be analyzed as incorporated into the verb, as they are modified by a NP in one of the usual NP modification constructions. The question of whether this construction type can be described as a N-V complex predicates is discussed again in Chapter VII.

The variation in the available evidence for the syntactic status of the deed element is common in the data, and is found in many construction types that contain three elements where the affected participant is not coded as a P argument. In these construction types, there is no other candidate for a P argument: the affected participant is coded as an oblique object, adnominal modifier, or a complement clause, and the doer participant is coded as an S/A argument. Thus, the only available candidate for P argument status is the deed element. However, as these deed elements are generally not flagged by DOM, short of evidence from ergative constructions, there is no positive support for the argument that they have P argument status. This situation is very similar to what is usually called “noun stripping” in descriptions and analyses of Native American languages (e.g., Miner 1986), and is discussed further in Chapter VII.

This construction type, in which the affected participant is coded as an adnominal modifier of the deed element, is also found in Iranian languages, such as Early Judeo-Persian and Sivandi, as is shown in (4.24a-b).

(4.24a) *šrx=y* *o* *ʔydr* *by-krd*
interpretation=LNK 3SG here PRV-do.PST.3SG
'he interpreted it here' (Early Judeo Persian, Gindin 2008, 3:14)

(4.24b) *rafīq=em* *niye=y* *kowgā=š* *kerd*
friend=1SG glance=LNK partridge=3SG do.PST
'my friend saw the partridge' (Sivandi, Lecoq 1979:129)

It should be noted that adnominal modification by nouns or NPs is coded across Iranian by the use of the Ezafe linker *=y*, seen in (4.24a-b). This marker, often a clitic on the head NP, is usually found between two NPs when the second modifies the first. The writing system of Early Judeo-Persian did not always provide an overt graphic representation of the Ezafe Linker (which was however, pronounced), much like the situation in New Persian (see Gindin 2008 for details). When the Ezafe Linker is not graphically represented, the result looks like two consecutive, juxtaposed NPs. Such sequences would have been treated in this study as instances of the main three-elements-construction type. Thus, it is possible that the low frequency of this construction type in Early Judeo-Persian is due to the lack of overt graphic representation of some grammatical markers, rather than to its actual rarity.

To summarize this subsection, the syntactic status of the deed element in this construction type varies across the languages analyzed here. In all of them, the deed

element NP always occurs right-adjacent to the verb, and is never flagged by DOM. This right-adjacent position of the deed element, however, is not a fool-proof test for P argument status. For example, in the main three-elements-construction type, the affected participant, not the deed element, is coded as a P argument, but the deed element, not the affected participant, is often right-adjacent to the verb.

One type of syntactic configuration which does provide positive evidence for the coding of deed elements as (non-prototypical) P arguments is found in languages with ergative alignment, at least in some corner of their grammar. Here, the deed element is indexed on the verb, as expected of a P argument. Such evidence is not available in languages without an ergative construction.

Thus, there is some crosslinguistic variation in the status of deed elements in this construction type: in some languages, deed elements are coded as P arguments, while in others there is no clear evidence for this coding, and it is difficult to interpret whether the deed elements are coded as (non-prototypical) P arguments or as predicative complements.

Finally, the coding of adnominal modifiers in Indo-Iranian is typologically diverse in terms of word order (e.g., head-linker-modifier in Sivandi and Early Judeo-Persian and modifier-GEN-head in Hindi and Nagamese), the morphosyntactic nature of the modification marker (e.g., old genitive case in Sanskrit and Pali, genitive postposition in Hindi and Nagamese, and the ezafe in Iranian), and other variables (see Plank 1995 for double case in Hindi genitive modifiers). Without the use of comparative concepts (such as adnominal modification) to characterize the syntactic status of the different

participants, this construction type would need to be artificially broken down into at least three distinct construction types.

4.2.3.3 A Argument, Deed Element, and a Comitative-Flagged Object

In this construction type, "do/make" is accompanied by three elements: the doer and the affected participants and the deed element. Doer participants are coded as A arguments, affected participants are coded as oblique objects flagged by a comitative marker, and the syntactic status of deed elements is crosslinguistically variable in much the same way as in the construction type described above in subsection 4.2.3.2.

This construction type, illustrated in (4.25a-c), often expresses reciprocal event types such as fighting or arguing. It is relatively common, found in many of the languages in the sample. In (4.25a), from Nagamese, the comitative post-position *ləgot* 'with' follows *bak* 'tiger'. The deed element is expressed by *judo* 'fight' and the doer participant is not overtly coded in the clause. In (4.25b), from Middle Persian, the doer participant is the head of the relative clause in which it is not overtly expressed by a NP, but is indexed on the verb. The affected participant in this clause is flagged by the preposition *abāg* 'with'. The deed element is expressed by the morphologically unflagged NP *pašn* 'contract, agreement'. In (4.25c), from Hindi, the affected participant is flagged by the comitative (and instrumental) marker =*se* and the deed element is expressed by the unflagged NP *bahs* 'dispute, argument'. The doer participant in (4.25c) is expressed by the question marker *kaun* 'who'.

(4.25a) *bak ləgot judo kor-i bak mar-ise.*
tiger COM fight do-NF tiger die-PST
'he fought with the tiger and killed it' (Nagamese, Sreedhar 1985:194)

(4.25b) *mard-ēw ke abāg zan ī xwēš pašn kun-ēd*
man-INDEF REL with woman LNK REFL contract do.PRS-3SG
'a man who makes a contract with his wife' (Middle Persian, RAF Q62)

(4.25c) *ab tum=se bahs kaun kar-e bhai!*
now 2SG=COM dispute who do-3SG Brother!
'now who will argue with you, brother?' (Hindi, Premchand 2017[1936]: 3.22)

The events expressed by this construction type cannot always be analyzed as having reciprocal semantics. In Hindi and Nagamese, for example, there are some instances where this construction type expresses events in which the doer participant initiates an event which affects the affected participant (often not physically). This is illustrated by (4.26a-b).

(4.26a) *kyõ kisii=se c^{h}ala-kapat kar-um*
why person=COM cheat-fraud do.PRS-1SG
'why would I cheat anyone?' (Hindi, Premchand 2017[1936]: 4.35)

(4.26b) *Itu nimite Jonah Isor logote parthana kor-ise*
DEM for Jonah God COM prayer do-PST
'and because of this, Jonah prayed to God' (Nagamese, Jonah 1:14)

4.2.3.4 A Argument, Deed Element, and a Complement Clause

In this construction type, "do/make" is also accompanied by three elements. Doer participants are coded as A arguments, the status of deed elements is yet again crosslinguistically variable as in the construction types described in subsections 4.2.3.2-3, and the third element is a complement clause. This construction type expresses events of perception, cognition, and utterance (PCU events) and phasal Aktionsart-type propositions such as "begin" and "finish". In (4.27a), from Ziyarat Mazandarani, the deed element is *fekr* 'thought', the doer participant is not overtly mentioned in the clause, but is indexed on the verb, and the verb is followed by a complement clause. In (4.27b) the deed element is *et* 'memory', preceded by a complement clause, which is a semi-indirect quote.

(4.27a) *fekr* *na-kard-en* *ke*
thought NEG-do.PST-3PL COMP
lebās=re *az* *tan=ešān* *darbiyār-en.*
clothes=OBJ from body=3PL take.off.PRS-3PL
'they did not think to take the shirts of their bodies' (Ziyarati, Shokri et al. 2013:225)

(4.27b) *mui* *ping-la-ʈa* *boli* *et* *ko-la*
1SG throw-PST-NMZ QUOT memory do-PST.3SG
'he remembered 'those are the diamonds I through away!' (Kotia Oriya, Gustafsson 1973a:121)

4.2.3.5 A Argument, Deed Element, and an R Argument

In this construction type, "do/make" is once again accompanied by three elements. Doer participants are coded as A arguments, affected participants are coded as R arguments, and deed elements are coded in a the same cross-linguistically variable manner as in the three previous construction types. This construction type is often used to express transfer or assistance events.

In several languages in the sample, such as Kupia and Bactrian, R arguments and P arguments are coded in an identical way at least in some constructions (see Sims-Williams 2010 for a survey of split ditransitive alignment in Bactrian). Instances where R and P arguments are not distinguished, are treated in here as instances of the main construction type with three elements, and therefore are not included here.

A difference in the coding of R and P arguments in Indo-Iranian is found primarily when R arguments are coded as oblique objects as in (4.28a-b). In (4.28a), from Middle Persian, the affected participant is flagged by the preposition *ō* 'to' and in (4.28b), from Sivandi, by the preposition *ba* 'to'. Both these prepositions are used to flag recipients with "give" and other transfer verbs.

(4.28a) *ke dāšn ō kas kun-ēd*
REL gift to someone do.PRS-3SG
'he who gives a gift to someone' (Middle Persian, DK6 23.1)

(4.28b) *ke ba hame komak kar-e.*
REL to 1PL help do.PRS-3SG
'(there is no one in this city) who will help us' (Sivandi, Lecoq 1979:135)

4.2.3.6 Caused (Change of) Location Construction Type

In this construction type also, "do/make" is accompanied by three elements, the doer participant, the theme participant and the location (or the ground). Doer participants are coded as A arguments, theme participants are coded as P arguments, and location participants are coded by locative expressions, often a NP flagged by a locative adpositions. The construction type, shown in (4.29a-b) from Gorani and Darai, expresses a caused change of location: the doer participant is responsible for the (re)location of the theme participant into some specified location.

In (4.29a), from Darai, the doer participant, which is not overtly expressed in the clause, puts the curry in the pot. The theme participant, the curry, is coded as a P argument *bʰat sag* 'curry (lit. cooked.rice greens)' and the location is expressed by the postpositional phrase *bogi=jə̃* 'in the pot'. In (4.29b), from Gorani, the doer participant is not overtly expressed in the clause, but is indexed on the verb form. The theme participant is expressed by the P argument *mēz* 'urine' and the location is coded by the post-verbal phrase *ī āw* 'this water'. This relative word order is common in Gorani and many other new Central Iranian languages. The verb is followed by the Gorani Directional Clitic.

(4.29a) *əse bogi=jə̃ bʰat sag kər-ikun*
then pot=LOC rice greens do-SEQ
'having kept the curry in the pot' (Darai, Dhakal 2013:23)

(4.29b) *mēz* *ma-kar-ām=a* *ī* *āw*
urine IND-do.PRS-1PL=DRCT DEM water
'we will urinate into the water (lit. we will do urine into the water)' (Gorani, Mahmoudveysi et al. 2012, 2:11)

The function of none of the elements in this construction type is comparable to that of a deed element in other construction types. That is, the change of location semantics of these clauses is not the result of the interpretation of the type of event expressed by some deed-type element, but instead is a result of the construction as a whole. Had clauses which belong to this construction type been accompanied by a deed element, they would fall within the set of construction types accompanied by four nominal elements.

4.2.3.7 A Argument, Deed Element, Locational Affected Participant

This section describes four distinct construction types which are infrequent in the data but are similar in terms in overall function and structure. In these construction types, the doer participant is coded as an A argument, and the affected participant is coded by a locative adpositional phrase originally expressing different spatial relations: "in", "on", "to, towards", and "from". The syntactic status of the deed element varies in much of the same way as in the construction types described in subsections 4.2.3.2-4. The main differences between this construction type and the change of location construction type described above is (a) the presence of a deed type element in this construction type and (b) the extension of the function of locative markers to flag the affected participant, and not necessarily the figure in a motion event.

These construction types express many different types of events, including motion and events of metaphorical transfer. In (4.30a-b) the affected participant is flagged by a preposition. In (4.30a), from Sivandi, it is *berey* 'on, upon' and in (4.30b), from Middle Persian, it is *pad* 'at, in, by'. The event expressed in (4.30a) is a metaphorical event of transfer, and the deed element is coded by *ta'rīf* 'awareness'. In (4.30b), the deed element is *xwārīh* (roughly) 'comfort'.

(4.30a) *ī* *berey kākā-gar=eš* *ta'rīf=eš* *me-kerd*
DEM on brother-PL=3SG inform=3SG IMPR-do.PST
'He informed his brothers' (Sivandi, Lecoq 1979:159)

(4.30b) *dēn* *ān* *baw-ēd* *ke* *pad harw dām* *xwārīh* *kun-ēd*
religion DEM be.PRS-3SG REL in all creature comfort do.PRS-3SG
'Religion is this: one who causes comfort to every creature' (Middle Persian, DK6 36)

In clauses in that express motion, the locative oblique argument actually carries locative semantics. This can be seen in (4.31), where the goal of motion is expressed by *bin des-e* 'to/in another land'. Thus, the function of *bin des-e* seems similar to that of the locative element in the construction type expressing caused change of location described above. However, in (4.31), the type of event expressed by the clause is expressed by the deed element *basa* 'movement'.

(4.31) *bin* *des-e* *basa* *ko-lai*
other land-LOC movement do-PST.3PL
'they went to another land' (Kotia Oriya, Gustafsson 1973a:201)

4.2.4 "Do/Make" with Four Nominal Elements

In many languages in the data set "do/make" is also found, albeit infrequently, together with four elements. Such clauses can be divided into two construction types. The first are mono-clausal examples which express a transfer event together with a change of state event. In such clauses, the four nominal elements are the doer participant, the theme participant, the recipient participant and a NP expressing the new state of the theme. The doer participant is coded as an A argument, the theme is coded as a P argument, the recipient is coded as an R argument, and the new state of the theme is expressed by a unflagged NP. The second type involves events of utterance where the addressee is overtly mentioned. In such examples the four elements the verb is associated with are the doer participant which is coded as an A argument, the addressee participant which is coded as an R argument, the deed element which is coded in a crosslinguistically variable manner as described in subsections 4.2.3.2-4 above, and a complement clause expressing the contents of the speech event.

In example (4.32a), from Middle Persian, a man gives his property to another as surety for a loan. Thus, (4.32a) involves both a (caused) change of state (the property becomes surety) and a transfer (the property becomes surety when given to the other man). Literally, the doer participant makes his property surety to another man. Example (4.32b), from Sivandi, involves an utterance event, the asking of a question, where the contents of the question asked are expressed by a complement clause which follows the main clause, and the addressee is expressed by a prepositional phrase with *ba* 'to'.

(4.32a) *mard-ēw ke xwāstag pad mard-ēw grawgān kun-ēd*
man-INDEF REL property in man-INDEF surety do.PRS-3SG
'a man, who gives (his) property to (another) man as surety (and takes a loan)'
(Middle Persian, RAF Q68)

(4.32a) *ye rū=i ba mehtar sohāl=em kerd ke ...*
one day=INDEF to chief question=1SG do.PST COMP …
'One day, I asked the chief …' (Sivandi, Lecoq 1979:113)

4.3 "Do/Make" as a Non-Main Verb

In at least three instances in the data, "do/make" is used as an auxiliary verb. This subsection gives a brief presentation of one such usage, and presents a slightly more detailed sketch of a second.

4.3.1 "Potentialis" in Sogdian (and Other Iranian Languages)

In this construction type "do/make" is used as an auxiliary verb which expresses ability or potential action. This construction type, illustrated in (4.33), is rare in the data, and is found only in Sogdian. Here, "do/make" is preceded by a non-finite (i.e., participle) form of the main lexical verb, which codes the type of event expressed in the clause and determines the number and form of arguments.

(4.33) *rty nwykr p'zn 'kw 'nwtr'y'n.sm'yk'smpwD'y L' 'wst'yt wnty*
and now mind to utmost.perfect.enlightenment NEG put.PTCP do.PRS.3SG
'then he cannot put his mind to the utmost perfect enlightenment' (Sogdian, MacKenzie 1976, VIM:23)

This construction type is not found in any of the Indo-Aryan languages in the data although Sims-Williams 2008 argues for a related construction in Sanskrit. However, there are some extinct and living Iranian languages, including Old Persian and Parthian (Korn 2013:33-37) where this construction type was either retained or some relics of it were preserved. Balochi languages, or dialects, have retained this construction type, as (4.34) below demonstrates.

(4.34) *man bi tīā* *āt-ag* *na-kurt-un*
1SG to 2SG.LOC come.PST-PTCP NEG-do.PST-1SG
'I could not come to you' (Turkmen Balochi, Axenov 2006:223)

4.3.2 "Dependent-Yet-Related" in Odian Languages (Kupia, Kotia Oriya)

This construction type is found in Kupia and Kotia Oriya, both Eastern Indo-Aryan languages of the Odian subbranch. "Do/make" is preceded by the non-final form of the main lexical verb, which can be described as a converb or a gerund, that denotes the lexical contents of the clause and functions. It is suggested here that the function of these constructions is to express that the events two clauses, the clause with this construction and the following clause, express thematically closely related, yet completely distinct events. That is, both are a part of the main event line of the narrative, or both express background information of the same type. This function is very similar to that of the Tamil (Dravidian) auxiliary construction with *viṭu* "leave" (Herring 1988), and as both Kotia Oriya and Kupia are spoken in close contact with Dravidian (especially Telugu), this construction type can perhaps be considered an areal phenomenon.

This construction type is found at least once or twice per page of glossed and translated text in both Kupia and Kotia Oriya. In the vast majority of instances, both in texts and in grammatical description (where examples are often elicited), the auxiliary "do/make" verb is also found in the same non-final form as the main lexical verb. The two constructions are schematically described in Table 4.1, and demonstrated in examples (4.35a-b).

Table 4.1 A scheme of "do/make" auxiliary construction in Kupia and Kotia Oriya

Kupia	Rest of clause (mostly arguments of V1)	V1-*a*	*ker-*
Kotia Oriya		V1-*i*	*ker-*
GLOSS		V1-NF	do-

(4.35a) *wa:gu=ka dek-a ker-a*
tiger=OBJ see-NF do-NF
'I saw the tiger…' (Kupia, Christmas & Christmas 1973a:39)

(4.35b) *goṭek pul duar-e rop-i kor-i*
INDEF flower yard-LOC plant-NF do-NF
'He planted a flower in the yard' (Kotia Oriya, Gustafsson 1973a:158)

In all languages in the sample, "do/make" is used in N-V complex predicate constructions, where the situation the clause is expressed is encoded by a non-verbal form in a preverbal position. Examples for such usages were amply provided above in subsection 4.2.2, and a further two are provided in (4.36a-b) below. In (4.36a), from Kupia, the deed element is *mo:sim* 'deception' and in (4.36b), from Kotia Oriya, it is *zuid* 'war'.

(4.36a) *he i:njo inayiŋ=ka mo:sim ker-l-o*
ha DEM 3PL=OBJ deception do-PST-3MSG
'Ha! He has deceived them' (Kupia, Christmas & Christmas 1973a:71)

(4.36b) *goʈek din-e se pila ar goʈek des=ke zuid kor-i*
INDEF day-LOC DEM child again INDEF country=OBJ war do-NF
'One day, this child again waged war on another country' (Kotia Oriya, Gustafsson 1973a:199)

The main argument against the analysis of examples like (4.35a-b) as instances of N-V complex predicates is the interaction of "do/make" with auxiliary verbs. The auxiliary verbs *gel* 'to complete' in Kupia expressing completive / perfective aspect, and *roi* 'remain' in Kotia Oriya expressing imperfective aspect, both follow the main lexical verb, as in (4.37a-b), where the main verbal predicate is "do/make". In contrast, when "do/make" functions as an auxiliary verb, and is combined with these further auxiliaries, it follows them, as shown (4.38a-b).

(4.37a) *jo wa:gu iŋ=ka ku:ɖanu ca:la ca:la kammo ker-a gel-ede.*
DEM tiger DEM=OBJ indeed very very work do-NF complete-FUT.3SG
'the tiger will do much more than this!' (Kupia, Christmas & Christmas 1973a:43)

(4.37b) *raza maizi=ke goʈek pila zonom kor-i roi-la*
king wife=OBJ INDEF child birth do-NF remain-PST.3SG
'meantime, the king's wife had given birth to a child' (Kotia Oriya, Gustafsson 1973a:188)

(4.38a) *goppa biḍiki der-an-a gel-a ker-a*
much terror bear-REFL-NF complete-NF do-NF
'I became terrified' (Kupia, Christmas & Christmas 1973a:41)

(4.38b) *mor-i roi kor-i uʈ-li*
die-NF remain do-NF get.up-PST.1SG
'having been unconscious (for some time), I got up' (Kotia Oriya, Gustafsson 1973a:171)

In the rest of this subsection, I will briefly attempt to describe the function of this construction type. This construction type is not used to express TAM semantics, shown by clauses where there are distinct TAM interpretations available for this auxiliary construction without additional grammatical marking. The examples provided for this construction type so far express events which occurred prior to the moment of speech, and could be interpreted as having perfective semantics. They all express events on the main event line of narratives. In contrast, examples (4.39-4.41) below show that this construction type can have a future temporal interpretation (events which occur after the time of speech), and can also express imperfective semantics.

(4.39) *mantrikumar-a a:ŋwu aranyawa:sum gec-a ker-a*
chief.minister-VOC 1SG pilgrimage go-NF do-NF
aranyum-te tappasu ker-inde
mountain-DAT penance do-FUT.1SG
'hey chief minister, I am going on a pilgrimage to the mountain to perform penance' (Kupia, Christmas & Christmas 1973a:47)

(4.40) *an-i* *kor-i* *maizi=ke* *ku-ai* *zibon* *kor-bi*
bring-NF do-NF wife=OBJ eat-CAUS life do-FUT.1SG
'(I will go to town and take some water), bring it, water the woman, and revive her' (Kotia Oriya, Gustafsson 1973a:198)

(4.41) *a:ji* *cedoyi* *ta-a* *ker-a* *ino=ka* *der-uka* *as-e*
today awake remain-NF do-NF DEM=OBJ carry-INF be.PRS-3SG
'(she said to herself), "Today, I will stay awake and catch him (in the act)."'
(Kupia, Christmas & Christmas 1973a:83)

Christmas & Christmas (1973b:305) describe the function of this construction in Kupia as coding what they call "Sequentive Aspect," and argue that its function is to code an event which "is completed before the action or state of the following clause takes place". Similarly, Gustafsson (1973b) describes the function of this Kotia Oriya construction as expressing sequential action. Both these analyses clearly hold in (4.42a-b) below, and in other examples presented in this subsection. In example (4.42a) the person first climbs the tree, and then shakes it in an attempt to make mangoes fall. In (4.42b), the speaker first consumes the medicine and then their health improves.[35]

(4.42a) *weg-a* *ker-a* *ambo* *i:s-e* *dunow-ite* *as-i*
climb-NF do-NF mango such-LOC shake-PRS.PTCP be-PRS.1SG
'I climbed the mango tree and I am shaking it' (Kupia, Christmas & Christmas 1973a:38)

[35] In their grammar of Standard Oriya, Neukom & Patnaik (2013) consider the corresponding form to be a "formal register" form of the general converb, expressing the temporal relationship between two events as sequential (as in (4.42a-b) above), or simultaneous (see below). They note that there is some difference between the converb followed by *kor-i* and converb forms not followed by *kor-i*.

(4.42b) *mui oso kai kor-i bol oi-li*
1SG medicine eat[36] do-NF good become-PST.1SG
'I ate medicine and became well' (Kotia Oriya, Gustafsson 1973b)

While the sequential analysis holds for many, and probably even most examples in the data, there are some crucial counterexamples, two of which are given in (4.43–4.44). In both, the clause with the "do/make" auxiliary is followed by a clause which is not temporally sequential to it. In (4.43), from Kupia, the tiger was fearsome and big while the speaker saw the tiger, and in (4.44), from Kotia Oriya, the marital status of the speaker and his opinion about the speech of wives are not sequential either.

(4.43) (a) *taruwata i:nja te:r bo:da adduru ja-a ge-t-ay*
then DEM female child shaking become-NF go-PRS-3SG
(b) *wa:gu=ka dek-a ker-a*
tiger=OBJ see-NF do-NF
(c) *goppa bayaŋka:rumuayina wa:gu well-o wa:gu;*
very ferocious tiger big-M tiger
'(a) then this woman starts shaking, (b) seeing the tiger. (c) it's a very fearsome tiger, a big tiger' (Kupia, Christmas & Christmas 1973a:39-40)

(4.44) *ce, mui goʈek munos oi kor-i,*
ha 1SG INDEF husband become do-NF
maizi-r kata=ke sun-bar bol nai
wife-GEN talk=OBJ hear-INF good NEG
'(he said) "ha! I'm a husband! It's not good to listen to wives talk!"' (Kotia Oriya, Gustafsson 1973a:142)

[36] The final *–i* is not overtly expressed when the verbal stem ends in a vowel.

Another situation in which this construction does not fit neatly with the sequential analysis can be seen in (4.45). The main lexical verb is the negative copular verb, used in a predicative possession construction. However, the state of not having suitable medicine is not over before the onset of the suffering described in the following clause.

(4.45) *sariyayina o:st-alu am-ci tenu nenj-a ker-a*
fitting medicine-PL 1PL-GEN with not.be-NF do-NF
ca:la ba:da-lu kallug-upa ja-te ta-wula
very pain-PL pervade-STAT come-PRS.PTCP be-FUT.3PL
'since we don't have fitting medicine, much suffering will be their lot' (Kupia, Christmas & Christmas 1973a:61)

To summarize, the examples given in this subsection show that the "do/make" auxiliary construction in Kotia Oriya and Kupia is not used to express TAM semantics. It has also been argued here that while a sequential interpretation is common in this construction, i.e., that the events described by this construction occur before the events in the following clause, such interpretations are not always accurate.

As suggested above, the main function of this "do/make" auxiliary construction type is to express the fact that two events are thematically related but distinct: that they express distinct events which are a part of the same "chunk" of discourse, in the sense that both are a part of the main event line of the narrative, or both express the same type of background information. If this is true, then when the two clauses are a part of the main event line of a narrative, they would naturally have a sequential temporal relationship.

To illustrate, consider examples like (4.46a-b), where the simple non-final form is not followed by the "do/make" auxiliary. In such clauses, the construction can be interpreted as a serial verb construction as in (4.46a), where people go and stay near the mango trees. In (4.46b), the crocodile did not finish looking for his brother, and the search would indeed continue. The crocodile's arrival at the grove occurs during, or as a part, of his search for his brother. Thus, the two are not distinct events.

(4.46a) *ku:ɖanu ʈo:nʈa-lu=te ma:nsu ta-a ge-t-ati*
indeed fruit-PL=LOC people stay-NF go-PRS-3PL
'(because they love mangoes) all the people go and stay among the fruit' (Kupia, Christmas & Christmas 1973a:43)

(4.46b) *se kumbir koz-i koz-i se ʈonʈa pali ai-la*
DEM crocodile search-NF search-NF DEM grove edge come-PST.3SG
'The crocodile search (for his brother) until he came to that grove' (Kotia Oriya, Gustafsson 1973a:143)

Compare these to (4.47a-b), where the "do/make" auxiliary is used. In (4.47a) the tapping of a tree is completed before the extraction of the juice begins. The two are not, and cannot be, simultaneous. In (4.47b), the act of climbing is done before the act of praying begins. That is, the child does not pray while in the act of climbing up the tree.

(4.47a) *e:ku ceʈʈu sind-a ker-a otta tinto pa:nukumu kaɖ-ula*
one tree skin-NF do-NF there from juice take-FUT.3PL
'they will tap (skin) a tree, and take juice from it' (Kupia, Christmas & Christmas 1973a:23)

(4.47b) *se pila goʈek goc-e sog-i kor-i durga=ke sumorna ko-la*
DEM child INDEF tree-LOC climb-NF do-NF Durga=OBJ prayer do-PST.3SG
'(at midnight) that child climbed up the tree and said a prayer to Durga' (Kotia Oriya, Gustafsson 1973a:122-123)

This subsection argued (a) that "do/make" is used in Kupia and Kotia Oriya as an auxiliary verb, and (b) that the function of the auxiliary constructions in these two languages is to express high thematic continuity, yet independence of the event or state described in the following clause. When these constructions express events on the main event line of narratives, the interpretation is of sequential events, which is probably the reason that previous analyses of these constructions described their function as encoding sequential semantics. When these constructions are used to express background events or states, the interpretation is not always sequential.

This is obviously simply a sketch of the function of this auxiliary construction. It does not discuss in detail its interaction with other auxiliary verbs, its interaction with the Aktionsart of the event expressed by the main clause, or its interaction with perfective aspect (see Hook 1991 for an overview of the usage of similar constructions in Indo-Aryan). Finally, as noted above, the function of expressing that two clauses are thematically related yet distinct events is also found in other non-Indo-Aryan languages of South / South Eastern India. For example, Herring 1991 argues that the function of the Tamil auxiliary *viʈu* "leave" is very similar to the one argued for here for "do/make" auxiliaries.

CHAPTER V

CONSTRUCTION TYPES OF "BE/BECOME" COGNATE VERBS

5.1 Preliminaries

This chapter describes and analyzes the construction types attested with cognate "be, become" verbs in the sample of Indo-Iranian languages studied here. Like the analysis of "do/make" presented in the previous chapter, the uses of "be/become" verbs are presented in two sets of construction types. The first set includes construction types in which "be/become" verbs function as the main verbal predicate of the clause; the second set includes patterns in which "be/become" verbs function as auxiliary verbs in periphrastic verbal constructions.

In contrast to "do/make" and "give", there are several different "be/become" verbs across Indo-Iranian. These verbs co-exist in many of the languages sampled here, either in suppletive patterns (e.g., expressing different tenses) or in different semantic niches (e.g., one "be/become" verb acquires dynamic semantics in some constructions). Two of the "be/become" verbs surveyed have clear copular cognates in both Iranian and Indo-Aryan languages. The first of these is the set of cognates of PIE *$*b^heuH_2$-* or *$*b^heH_2u$* (Pokorny 1959:146), and the second is the set of cognates of PIE *$*H_1es$-* (Pokorny 1959:340). Cognates of these two verbs have been found, to some degree, in all languages surveyed. They are complemented by two other verbs which function as copulas in only one of the main branches of Indo-Iranian. The first, found as *(a)ch-* or *as-* in many Indo-Aryan languages, is argued by Masica (1991:285) to be a cognate of the Sanskrit root *kṣi-* "to dwell, live", which is a cognate of PIE *$*tḱei$* (Pokorny 1959). Iranian cognates of this verb in the data do not function as "be/become" verbs. The

second verb is reconstructed in Proto-Iranian as **maH* in Cheung 2007:257, who cites several etymological hypotheses. Cheung's hypotheses are sharply criticized in Yakubovich 2010, who argues in favor of a PIE etymology of "grow" with cognates in Anatolian (as a finite verb), and Slavic and Italic (as an adjective). Cognates of **maH* were not found in the Indo-Aryan data.

Indo-Iranian cognates of these four PIE verbs are still found in some (further) suppletive relationship with additional verbs from other sources (e.g., Hindi *t^{h}a-* "be/become" past tense form). Instances of such additional verbs were also collected, but construction types in which only these additional verbs were used are not included here. Constructions where only these additional verbs are attested mostly belong to construction types in which the "be/become" functions as an auxiliary (e.g., the Hindi Past Tense construction). As for construction types in which "be/become" functions as the main verbal predicate, there was little to no difference in the distribution of the main four "be/become" verbs and additional suppletive verbs in the data.

5.2 "Be/Become" as Main Verbal Predicate

This subsection describes and analyses the construction types in which cognate "be/become" verbs function as the main verbal predicate. These construction types include clauses expressing all six nominal predication[37] functional domains identified in the typological literature (e.g., Stassen 1997, Payne 1997): equation (identification), predicate attribute, proper inclusion, predicative possession, the predicate locative, and

[37] I differentiate between nominal predicate, which is an instance of a nominal element functioning as the predicate of a clause and nominal predication, which is a set of six (possibly seven) functional domains often encoded by clause constructions with a nominal predicate (either with or without a copular verb).

the existential. Additional construction types in which "be/become" verbs function as the main verbal predicate resemble N-V complex predicate constructions.

The definitions of the six nominal predication functional domains used in this dissertation in general and in this chapter in particular follow Payne 1997:111-113. Clauses encoding the *equative-identificational* (henceforth *equation*) functional domain encode that the referents of two phrases in a clause are identical. Clauses encoding the *predicate attribute* (henceforth *predicate property*) function are clauses in which one phrase expresses a property (i.e., a characteristic) of the referent of another phrase. As Payne 1997:111 notes, in many languages, including some of those analyzed for this dissertation, adjectives often function as the predicate in clauses expressing this function. There are, however, many instances in which the predicative phrase is not a lexical adjective. Examples include English *I am an asshole* and *she is a genius*. Thus, I prefer the term *predicate property* to avoid necessarily alluding to adjectives in this function. Clauses encoding *proper inclusion* predicate the inclusion of a referent of one phrase in a set denoted by another phrase (e.g., *she is a linguist*, when referring to her profession).

These three functions (equation, predicate property, proper inclusion) are often confounded and called "nominal predicate" or "nominal predication" functions (see, e.g., Stassen's WALS entry in Stassen 2013a). Lumping these functions together is sometimes justified by their grammar in Indo-Iranian languages (see Chapter VIII for analysis and details).[38] For this reason, I refer to the set that includes these three functions as the *core nominal predication* functions.

---

[38] This is true for most of the usage of cognates of "be/become" verbs. It is less true for instances of verbs of motion and posture which often grammaticalize as copulas. There are instances, briefly reviewed in Chapter

The other three (typically) nominal predication functions are *predicative possession*, the *predicate locative*, and the *existential*. *Predicative possession* is defined here following Payne 2009 (who follows Langacker 1993) to include clauses encoding an intimate or control relationship between two referents (a reference point and a participant, in Langacker's 1993 terminology). Clauses expressing predicative possession thus include clauses expressing kin or familial relations between two entities (intimate relationship), part-whole relations (intimate relations), and ownership relationship (control relationship).

The *predicate locative* function is defined here, again following Payne 1997:112 (and also Creissels 2013, 2014), as predicating the location of some entity in a figure-ground relationship. The *existential* function is defined as predicating the existence of an entity in some location. These two functions differ in their configuration or construal of the relationship between the figure and the ground. In clauses expressing the predicate locative functional domain, the figure is given ("old") information, and the ground (i.e., the location of the figure) is new information. In clauses expressing the existential functional domain, the figure is (typically) not given information, and its existence (often in a specified location) is predicated. Several linguists have commented on the functional and grammatical relationship between those two functions (e.g., Clark 1978, Creissels 2014). The relationship of these two functional domains in terms of their grammatical encoding is examined in Chapter VIII.

VII, in which such verbs are used in constructions encoding predicate property or proper inclusion, but not equation. This is especially true in the earlier stages of the grammaticalization process.

The rest of this section presents the different "be/become" construction types where it is the only verbal predicate. Each construction type is defined based on a pairing of form and function. The form of these construction types is defined in terms of comparative concepts, similarly to the definitions used with "do/make". The functions are mostly the six nominal predicate functions.

5.2.1 "Be/Become" with Two Morphologically Unflagged NPs

Across the data used for this study, "be/become" is most often accompanied by two unflagged NPs. This is the most common configuration of coding means with which "be/become" is attested in the data in three senses. First, examples of this configuration of coding means are found in all sixteen languages analyzed here. Second, this the most common configuration of coding means in which "be/become" is the main verbal predicate in all sixteen languages combined.[39] Finally, this is a functionally versatile, or semantically un(der)marked, configuration of coding means. Thus, it is referred to here as the Main Pattern of "be/become". The main pattern is used across the data to express all six nominal predication functions. The following subsections describe its usage to encode the three core nominal predication functions, predicative possession, predicate locative, and existential functions.

[39] In some languages, including Middle Persian, Nagamese, and Hindi, the most frequent use of "be/become" is as an auxiliary verb.

5.2.1.1 Main Pattern Expressing Core Nominal Predication.

The Main Pattern of "be/become" occurs most commonly in clauses expressing the three core nominal predication functional domains. Initial examples are given in (5.1a-c). Example (5.1a), from Gorani, expresses equation; the referent of *īrij*, a proper name, is predicated to be identical to the referent of *kuřaš* 'his son'. Example (5.1b), from Palula, expresses a predicate property. The speaker, expressed by the 1SG independent pronoun *ma*, is said to be *na ʈing* 'not firm'. Example (5.1c), from Ziyarat Mazandarani, expresses proper inclusion. The husband of the speaker is said to be a Galesh, a specific type of herdsman, which is a culturally central concept for Ziyarat Mazandarani speakers (as well as other related ethnic groups).

(5.1a) *īrij kuř=aš biya*
irij son=3SG be.PST.3SG
'Irij was his son' (Gorani, Mahmoudveysi et al. 2012, 5:8)

(5.1b) *ma na ʈing bhí-l-u*
1SG.NOM NEG firm be-PST-MSG
'(I said) I was not firm' (Palula, Liljegren & Haider 2015:52-53)

(5.1c) *me āɣā gāleš bie*
1SG.GEN mister galesh be.PST.3SG
'My husband was a galesh (cowherd)' (Ziyarat Mazandarani, Shokri et al. 2013:87)

The coding of the two NPs accompanying "be/become" in this construction type is very similar across the sample languages. Morphologically, both the topic and the

predicate are unflagged NPs, as in (5.1a-c) above. In all languages where verbal indexing is available (at least in some constructions), the verbal copula in this construction type indexes the topic NP.

Across all languages in the sample, the topic NP in this construction type is coded as an S or S/A argument. When A and S arguments differ in their morphological flagging, the topic NPs in this construction type are almost never coded as A arguments. Instances in which topic NPs are flagged as A arguments are rarely found in Middle Persian, and can usually be explained by their highly accessible information status in those clauses (Shirtz 2016).

The topic NP behaves like an S argument when it comes to indexing on the verb. In all constructions where verbal indexing is found, the topic NP is indexed on the verb. This was shown in (5.1a-c) above, and is also illustrated in (5.2a-b) from Middle Persian. In Middle Persian, for example, while S arguments are indexed on verbs in both the past and present stems, P arguments are indexed only on verbs in the past stem and A arguments are indexed only on verbs in the present stem. Now, in (5.2a) the verb is in the past tense, and indexes the topic NP via an auxiliary verb. In (5.2b) the verb is in the non-past form, and again indexes the topic NP.

(5.2a) *arzānīg būd h-ēm*
worthy be.PST be.PRS-1SG
'I was worthy' (Middle Persian, AWN 4.14)

(5.2b) *nakkīrā baw-ēm*
denier be.PRS-1SG
'I will deny it (lit. I will be a denier)' (Middle Persian, DK6 59)

The predicate NP is not coded like any core argument. Across the sample, predicate NPs do not behave like S or A arguments, as they are never flagged by ergative markers, and they are never indexed on the verb. Furthermore, predicate NPs are not coded like P arguments, as they are never found flagged by DOM (in those languages which exhibit DOM), and they are never indexed on the verb in ergative constructions.

Predicate NPs, however, are bona fide NPs. They can be modified by clitic pronouns as in (5.3a), modified by adjectives, and conjoined with other predicate NPs as in (5.3b), or can be modified by genitive-flagged NPs as in (5.3c). Examples are given from three languages, but similar examples of this construction type are found in all sixteen languages of the sample.

(5.3a) *kākā=t hand-ām*
brother=2SG be.PRS-1SG
'I am your brother' (Sivandi, Lecoq 1979:160)

(5.3b) *u=šān gāh ud nišast abēr rōšn ud brāzāg ud purr.xwarrah būd*
and=3PL place and seat very light and shiny and full.glory be.PST.3SG
'and their place and seat was very bright, shining, and full of glory' (Middle Persian, AWN 7.2)

(5.3c) *ok^{h}ni-hi sor-kə rup b^{h}əi-te*
immediately-EMPH pig-GEN form be-TMP
'as she was in the form of a pig' (Darai, Dhakal 2013:94)

The examples given so far demonstrate that the predicate in this construction type is expressed by an unflagged NP, and it is not coded like any core argument. Thus, it is justified perhaps to refer to these NPs as "predicative complements".

This subsection only provided a sketch of this construction type in which cognate "be/become" verbs are accompanied by two unflagged NPs in clauses expressing the three core nominal predication functional domains. The sketch here has been limited mostly to the coding properties of the two NPs expressing the topic and the predicate. There are several other grammatical issues which are beyond the scope of this subsection. These include the interaction of this construction type with dynamic readings ("become") and its interaction with *h-* or *bʰ-* "be/become" verbs, the availability of non-indicative, especially imperative, semantics, and the interaction of this construction type with manner adverbs, negation, and other types markers and categories. These seem to be a major point of crosslinguistic variation in the sample, and cross linguistically in general (see, for example, Stassen 1997). An analysis which takes all these variables into account is well beyond the scope of this study.

5.2.1.2 Main Pattern Expressing the Predicate Locative and Existential

The coding properties of the construction type described in the previous subsection, 5.2.1.1, involve two unflagged NPs, encoding the topic and the predicate, and a "be/become" verb which indexes the topic NP. The construction type described in this subsection has the same configuration of coding means. The difference between the two construction types is that while construction type described in 5.2.1.1 expresses core nominal predication, the one described in this subsection encodes the predicate locative and the existential functional domains. I will refer to the two NPs in this construction type by their roles: *figure* NP refers to the entity being located, while *ground* NP refers to the background the figure is located in relation to.

The clearest instances of this construction type are clauses in which the ground element is expressed by an unflagged lexical NP which usually refers to a well known location. This is demonstrated in (5.4a-b) from Gorani and Kupia. In both examples, the ground is expressed by an unflagged NP, which denotes a place name: in (5.4a) by the NP *faransa* 'France' and in (5.4b) by the NP *kilagaḍa* "Kilagada" (a town in Andhra Pradesh state). In both Kupia and Gorani, a place name can be accompanied by locative adpositional markers when functioning as the predicate locative of a clause. This is shown in subsection 5.2.3 below.

(5.4a) *usā āsā faransa biya*
master then France be.PST.3SG
'At that time, the master was in France' (Gorani, Mahmoudveysi et al. 2012: 4:65)

(5.4b) *aŋgrar santa kilagaḍa as-e*
Tuesday market Kilagada be-PRS.3SG
'(there's) a Tuesday market in Kilagada' (Kupia, Christmas & Christmas 1973a:18)

The exact range of use of this construction type is a bit difficult to delimit, because of the grammaticalization processes in which lexical nouns are reinterpreted as locational, relational nouns, or eventually even grammaticalize as adpositions. The intermediate stages in this grammaticalization process, where specific instances can be interpreted as one or another item, pose analytic difficulties. I illustrate the grammaticalization cline by two Sivandi locational nouns: *jelow* 'in front, forward' and *sar* 'head, on top'. While the first element, *jelow*, can never be interpreted as a NP in

clauses encoding (stative) predicate location, the status of the second element, *sar* 'head, on top', is sometimes more ambiguous.

In the data, the Sivandi marker *jelow* 'in front, forward' expresses only stative locations and goals of motion. When it expresses a goal of motion it can be used unmodified, as in (5.5a), or it can be modified by a NP in the Sivandi Ezafe construction, as in (5.5b). The Sivandi Ezafe construction is the usual construction when a NP modifies another NP. When *jelow* expresses a stative location, however, it is always modified by a clitic pronoun, as in (5.5c) where *jelow* is modified by the third person singular clitic pronoun *=eš*.[40]

(5.5a) *ši* *jelow*
go.PST.3SG in.front
'(the fox ran), went forward' (Sivandi, Lecoq 1979:152)

(5.5b) *ši* *jelow=i* *kākā=š*
go.PST.3SG in.front=LNK brother=3SG
'he went in front of his brother' (Sivandi, Lecoq 1979:89)

(5.5c) *darye=y* *owe* *jelow=eš* *en*
sea=LNK water in.front=3SG be.PRS.3SG
'the lake of water is before you' (Sivandi, Lecoq 1979:112)

In examples (5.5a-c) above, *jelow* cannot be interpreted as a fully lexical NP which is modified by another NP or a clitic pronoun. Rather, its syntactic status is better interpreted as a relational noun. Thus, it is interpreted here as a directional marker

[40] Modification of *jelow* by a clitic pronoun is also found when it expresses a goal of motion.

relative to the addressee in (5.5a), and as a locative marker attached to "your brother" in (5.5b) or to the third person singular clitic pronoun in (5.5c).

While *jelow* only expresses locations in the data and is not used as a lexical noun, other elements with a similar function co-exist together with their lexical source. This makes the interpretation of some specific instances of their usage difficult, as they can be interpreted both as lexical nouns or as relational locative nouns. These instances can be thought of as the bridge contexts, between the original, lexical usage of these elements and their innovative usage as relational nouns.

One such element is *sar* 'head'. In (5.6a-b) it is used in a locative sense, and in both examples, it is modified by another NP in the Sivandi Ezafe construction. When used as a relational noun, *sar* 'head' has the meaning of "top" or "on top", as demonstrated in (5.6a), where a partridge goes to the top of a rock.

(5.6a) *ye kowgi ām-ey sar=i ye vard-i*
INDEF partridge go.PRS.3SG head=LNK INDEF rock-INDEF
'(he saw that) a partridge went to the top of a rock (lit. to the head of a rock)'
(Sivandi, Lecoq 1979:128)

In (5.6b) below, however, two interpretations of *sar* are available, and the clause does not provide any means for deciding between the two. The veil can be "on top" of the brother, or "on the head" of the brother. In the first interpretation, *sar* functions as a relational noun, and the clause is an instance of a construction type where "be/become" is accompanied by a NP and a locative element, and not an instance of the construction type described in this subsection. In the second interpretation, *sar* in (5.6b) is interpreted with

its lexical meaning of "head", and the clause is an instance of the construction type in which "be/become" expresses the predicate locative function. Where the meaning is ambiguous, I chose the literal, original or source, interpretation.

(5.6b) *čāder sar=i kākā=t bi, sar=i jom'a bi*
veil head=LNK brother=2SG be.PST.3SG head=LNK Friday be.PST.3SG
'The veil was (on) your brother's head, it was (on) the head of Joma!' (Sivandi, Lecoq 1979:89)

5.2.1.3 The Main Pattern Expressing Predicative Possession

In this construction type, the coding properties of the verbal copula and its associated NPs are identical to those found in the construction types described in subsections 5.2.1.1 and 5.2.1.2. In contrast, however, the clauses analyzed here express predicative possession. This construction type is illustrated in (5.7a-b) below, from Middle Persian and Darai.

In (5.7a), from Middle Persian, "be/become" is accompanied by two unflagged NPs: *harw kas* 'every person' and *ciš=ē* 'a thing'. It is obvious from both the content of the two NPs and from the following clause that the clause in question does not mean 'every person is a thing' but rather 'every person has a thing'. Similar examples are found throughout the text from which this example is taken.

Example (5.7b), from Darai, is a bit more cumbersome. In this example, "be/become" is accompanied by two unflagged NPs: 'his wife' and 'son'. Clearly, an interpretation of this clause as expressing the equation or proper inclusion makes no

sense. The only logical interpretation is a possessive one, i.e., that his wife did not have a son.[41]

(5.7a) *harw kas ciš=ē ast*
every person thing=INDEF be.PRS.3SG
ī=š az abārīg ciš āzarmīg-tar
LNK=3SG from other thing precious-CMPR
'Every person has one thing which is dearer than other things' (Middle Persian, DK6 33)

(5.7b) *ukʰ=rə dulhi=k cʰawa nidz=bʰəi-lə*
3SG=GEN wife=3SG.POSS son NEG=be-PST
'his wife did not have a son (lit. his wife was not a son)' (Darai, Dhakal 2013:79)

The expression of possession without a "have"-type verb but with two morphologically unmodified NPs is also found in other languages in the sample, although in them the clause is verbless, as illustrated by (5.8) from Kotia Oriya.

(5.8) *se raza besi sompoti*
DEM king much possessions
'the king had many possessions' (Kotia Oriya, Gustafsson 1973a:174)

[41] The usual expression of predicative possession with "be/become" in Darai follows Heine's 1997 Genitive schema ("X's Y exists"). It is demonstrated below.

5.2.2 "Be/Become" in Clauses Expressing Predicative Possession

Subsection 5.2.1.3 above showed the infrequent uses of "be/become" to express predicative possession with two unflagged NPs. Commonly, when "be/become" is used in clauses encoding predicative possession, the NP expressing the possessor is flagged by at least some morphological marker. One finds possessors coded as adnominal modifiers of the NP expressing the possessed, flagged by markers associated with human or animate goals (recipient and benefactive markers), locative markers, or comitative markers. This range of encoding of predicative possession shows that "be/become" predicative possession construction types in Indo-Iranian are a typologically diverse field. This diversity is likely correlated with the disintegration of the old Indo-Iranian case system, which forced speakers to innovate constructions expressing predicative possession.

5.2.2.1 Possessor as Adnominal Modifier

In this construction type, the possessor argument is coded as an adnominal modifier of the possessed argument. In clause constructions where an argument is indexed on the verb, it is the possessed argument which is indexed, in accordance with Heine's 1997 Genitive Schema "X's Y exists".

This construction type is common across the family, and is found in several of the older Indo-Iranian languages such as Sanskrit and Old Persian[42] as shown in (5.9a-b). In (5.9a) from Vedic Sanskrit, the possessor is flagged by the Genitive case and the

[42] While Old Persian was not included in the sample of languages for this dissertation, I include Old Persian examples here as they illustrate the effects of the (early) disintegration of the old case system on the expression of predicative possession.

possessed is flagged by the Nominative case, and is indexed on the verb. In (5.9b), from Old Persian, the possessor is flagged by the Genitive-Dative case, following the syncretism of the Proto-Indo-Iranian Genitive and the Dative case forms, and the possessed is in the Nominative case and is indexed on the verb.

(5.8b) *mánor ... ɾ̥ṣabʰá āsa*
manu.GEN … bull.NOM be.PST.3SG
'Manu had a bull' (MacDonnell 1958:320, glosses mine)

(5.9b) *dārayavahauʃ puçā aniyai=ciy āha(tā)*
darius.GEN.DAT son.NOM other=also be.3SG
'Darius had other sons too' (Old Persian, Xpf 28-29, glosses mine)

This construction type is not limited to old Indo-Iranian languages in which the old Genitive case had been retained. In examples (5.10a-c) below, from Darai, Sivandi, and Gorani, all Modern Indo-Iranian languages, the possessor is flagged as an adnominal modifier of the possessed NP. In (5.10a), from Darai, the possessor *hame* 'we' is flagged by the genitive marker *=rə* and functions as an adnominal modifier of the possessed *radza* 'king'. In (5.10b-c), from Sivandi and Gorani, the possessor is expressed by a clitic pronoun. The possessed is expressed by a morphologically unflagged NPs, "horses" and "one bag," in (5.10b-c) respectively. Thus, examples (5.10a-c) can all be interpreted as instances of Heine's Genitive Schema ("X's Y exists").

(5.10a) *e pəhila hame=rə radza nə=bʰə-i*
EXCL first 1PL=GEN king NEG=be-3SG
'eh, first we didn't have a king' (Darai, Dhakal 2013:65)

(5.10b) *usūr-ā=mā bi*

horse-PL=1PL be.PST.3SG

'we have horses' (Sivandi, Lecoq 1979:152)

(5.10c) *yak kaškūł=iš h-ē*

one bag=3SG be.PRS-3SG

'he has a bag' (Gorani, Mahmoudveysi et al. 2012, 4:32)

The analysis of these Sivandi and Gorani constructions is actually more complex than these two examples might suggest. Both constructions have undergone a similar process of reanalysis resulting in an expression of the possessor that is more similar to that of S/A arguments (see discussion in subsection 5.2.2.5, under "transitivization" per Stassen 2009, 2013b).

5.2.2.2 Possessor as Goal or Indirect Participant

In the construction type analyzed in this section, "be/become" is the main verbal predicate in clauses expressing predicative possession and the possessor is expressed as an animate goal or "indirect participant", i.e., recipient, benefactive, or external possessor (see definition and brief discussion in Haig 2008:58-61). The morphological status and phonological form of markers of indirect participation vary across Indo-Iranian, and they include a set second position clitic pronoun (in some Iranian languages) and a NP flagged by some specific marker (in Indo-Aryan and some Iranian languages).

Another locus of variation in the coding of indirect participants has to do with the range of semantic (micro)roles that is flagged by the same marker used to flag indirect participants. In some of the languages analyzed here, the functional range of these

markers is limited to indirect participants, while in others it includes core-arguments as well. Haig 2008 showed that across Western (and Central) Iranian, the use of the clitic pronoun to flag A arguments is innovative, and extends from its function of expressing indirect participants. Shirtz 2016 argues that the Middle Persian clitic pronouns were also extended to code S/A arguments in at least one clause construction. Thus the functional range of the clitic pronouns in Middle Persian includes both indirect participants and core arguments. In Kotia Oriya, for example, possessors in this construction type are flagged by the same markers as P and R arguments.

Across the Old and Middle Iranian languages possessors in this construction type are often expressed by a second position clitic pronoun. This is illustrated by examples (5.11a-c) from Old Persian, Bactrian, and Middle Persian, where the possessor is expressed by the second position clitic pronoun, and the possessed is expressed by an unflagged NP and is also indexed on the verb.[43]

(5.11a) *utā=taiy tauhmā vasiy biyā*
and=2SG.GEN.DAT seed much be.SUBJ.3SG
'may you have much seed!' (Old Persian Kent 1953 DB IV:75, cited in Haig 2008:28)

(5.11b) *asida=men χan spandagan girl-ind misid ast*
and=1PL house spandagan call.PRS-3PL TOP be.PRS.3SG
'and we have a house in that place they call Spandagan' (Bactrian, Sims-Williams 2001, V6)

[43] These second position clitic pronouns are cognates of the clitic pronouns used to express possessors in the Sivandi and Gorani examples in (5.10b-c). Following what is sometimes called "head attraction" (Haig 2008:336-338), the Sivandi and Gorani clitic pronouns got attached to their functional head.

(5.11c) *zan* *gīr* *tā=t* *frazand baw-ēd*
woman catch.IMPR.SG until=2SG child be.PRS-3SG
'Take a wife, so you'll have a child' (Middle Persian, PRDD 36)

The functional range of the second-position clitic pronouns in Old Persian, Middle Persian, and Bactrian is wide, and includes external possession, benefactive, as well as core argument functions, as illustrated in (5.12a-b) from Middle Persian. These examples show the use of the clitic pronoun to express different indirect participant functions. In (5.12a), the clitic pronoun expresses the external possessor, that is, the person whose hand is put in the oven. In (5.12b), the clitic pronoun expresses the benefactive, that is, the person for which the money is stolen.

(5.12a) *ud* *ka=š* *dast* *be andar ō* *tanūr bar-ēd*
and if=3SG hand to in to oven carry.PRS-3SG
'And if one puts his hand into the oven' (Middle Persian, PRDD 37a)

(5.12b) *pad-iz* *drahm ī=š* *stān-ēnd* *duz* *h-ēnd'*
by-FOC drahm LNK=3SG take.PRS-3PL thief be.PRS-3PL
'Indeed in the matter of the drahm they stole for him, they are thieves' (Middle Persian, PRDD 30).

The Middle Persian clitic pronoun is also used to express core arguments in a split alignment pattern: A arguments when the verb is in the past tense, and P arguments when it is in the present tense (see Haig 2008:112-117). Additionally, it expresses S/A arguments in the "need" auxiliary construction (Shirtz 2016). The use of the clitic pronoun to code core arguments depends on TAM, but its encoding of indirect

participants as in (5.12a-b), or possessors in the predicative possession construction is not limited to specific TAM semantics or specific verb forms. To illustrate, in (5.11c) above it expresses the possessor while the "be/become" verb is in the present tense, whereas in (5.13) below the clitic pronoun expresses the possessor in the Middle Persian Predicative Possession construction where the verb is in the past tense.

(5.13) *ke=šān pad gētīg stōr būd*
REL=3PL in world cattle be.PST.3SG
'(people) who in the material world had cattle' (Middle Persian, AWN 77.3)

Central to clause alignment patterns as they might be, the clitic pronouns in Old Persian, Middle Persian, and Bactrian are nevertheless pronominal. As such, they are not suitable for expressing a possessor argument (or any other referent, for that matter) as a first mention in discourse or when their referents need to be re-introduced to the discourse. Speakers solved this "problem" using different grammatical strategies. Some of these evoke the same "goal" schema, while others do not (see subsection 5.2.1.3 above, for example).

In Old Persian, where the old Indo-Iranian morphological case system was still viable (albeit more narrow), possessors were also expressed by lexical NPs in the Genitive-Dative case (see example (5.9b) above). In Middle Persian, predicative possession was expressed by "be/become" accompanied by two unflagged NPs, as illustrated above in (5.7a). Alongside this, one finds in Middle Persian a third possibility to express the possessor. In example (5.14) the possessor, *ōy wirāz* 'that Wiraz' is flagged by the postposition *rāy* 'to, for'. In Middle Persian, *rāy* is used to flag benefactives,

reasons, and, infrequently, recipients. In later forms of Persian, such as Standard New Persian or Early Judeo Persian, cognates of *rāy* have been also used to express core arguments, such as P arguments (as DOM) and R arguments (in some varieties, but not in Standard New Persian).

(5.14) *ud ōy wirāz rāy haft xwah būd h-ēnd*
and DEM wiraz to seven sister be.PST be.PRS-3PL
'and Wiraz had seven sisters' (Middle Persian, AWN 2.1)

In Bactrian, there is a small number of examples of clauses with "be/become" expressing predicative possession where the possessor is expressed by means other than the clitic pronoun. In example (5.15) below, taken from a contract, the first person pronouns refer not to the person writing the contract, but to the persons signing the contract. The possessors are expressed by a combination of the first person singular pronoun, in the Bactrian Oblique case form, followed by the proper name of the person signing the contract. The possessed is expressed by an unflagged NP which is also indexed on the verb. Another strategy used in Bactrian to express possessors is the companion strategy, reviewed in section (5.2.2.4) below.[44]

[44] Furthermore, Bactrian, unlike Old and Middle Persian, has also a grammaticalized possessive verb from "hold".

(5.15) *man pōtan od man miyar ... saxwan od*
1SG.OBL PN and 1SG.OBL PN word and
lad od aχas stad
lawsuit and dispute be.PRET.3SG
'I, Potan, and I, Miyar, had a quarrel and a dispute…' (Bactrian, Sims-Williams 2001, S10)

The expression of predicative possession by constructions that belong to Heine's Goal Schema is also found in some of the Indo-Aryan languages. Here, it is demonstrated by examples (5.16a-b), from colloquial Gujarati and Kotia Oriya. In both examples the possessor is flagged by the primary object marker, *=ne* in Gujarati and *=ke* in Kotia Oriya, and the possessed is expressed by an unflagged NP which is also indexed on the verb.

(5.16a) *raja=ne ek kûwəri ch-e*
king=OBJ one princess be-PRS.3SG
'the king had one princess' (Gujarati, Christian 1987:276)

(5.16b) *se dangɽ-i=ke pila oi ga-la*
DEM youngin-F=OBJ child be.NF go-PST.3SG
'The young girl had a child' (Kotia Oriya, Gustafsson 1973a:194)

The functional range of the Colloquial Gujarati *=ne* and the Kotia Oriya *=ke* markers includes both core arguments and indirect participants. Both markers are used to flag P arguments as shown in examples (5.17a-b), and R arguments as in (5.18a-b).

(5.17a) *marwad gujərat əne səurasʈrə=na məll.raj-o=ne həraw-ine*
marwad Gujarat and saurastra=GEN.MPL wrestler.king-MPL=OBJ defeat-CV
'having defeated the great wrestlers of Marwad, Gujarat, and Saurashra, (they came to Kutch)' (Gujarati, Christian 1987:206)

(5.17b) *ek dant rakias-ni=ke mar-i*
one tooth dragon-F=OBJ hit-NF
'he killed the one-toothed dragon, (and then married Iraboti)' (Kotia Oriya, Gustafsson 1973a:126)

(5.18a) *ek=nû nam hû ap=ne ap-û*
one=GEN.NSG name 1SG 2PL=OBJ give-PRS.1SG
'I will give you the name of one (person)' (Gujarati, Christian 1987:210)

(5.18b) *lokio ʈanka dangɽ-a-mon=ke de-la*
million rupee youngin-M-PL=OBJ give-PST.3SG
'He gave a million rupee to the young men' (Kotia Oriya, Gustafsson 1973a:117)

Apart from P and R arguments, both *=ne* and *=ke* are used to flag animate goals. In (5.19a), *=ne* is used for the first person in a metaphoric "fear" clause with the motion verb "enter". In (5.19b), *=ke* is similarly used to flag the metaphoric goal of a PCU event of the motion verb "come".

(5.19a) *mə=ne=to* d^h*ast-i* pet^h*-y-i*
1SG=OBJ=TOP fear-FSG enter-PTCP-FSG
'I was afraid (lit. fear entered me)' (Gujarati, Christian 1987:210)

(5.19b) *se din=e se pila=ke et ai-la*

DEM day=LOC DEM child=OBJ memory come-PST.3SG

'on that day, the child's memory came back' (Kotia Oriya, Gustafsson 1973a:121)

In summary, this subsection has described the "be/become" construction type in which "be/become" functions as the only verbal predicate in clauses expressing predicative possession, and the possessor is flagged as a human goal or indirect participant. Instances of this construction type belong to Heine's Goal Schema ("Y exists to/for X"), and are found across both Iranian and Indo-Aryan. The functional range of the possessor marker in these constructions is not identical across languages, but always includes indirect participation functions and human goals.

5.2.2.3 Possessor as (Stative) Location

The expression of predicative possession by clauses which follow Heine's 1997 Location Schema ("Y is located at X") is infrequent in the sample, and seems to be rare across Indo-Iranian in general. There are, however, several examples of clauses expressing predicative possession with "be/become" as the main verbal predicate, and the possessor flagged by some (synchronic) locative marker.

In (5.20a) below, from Colloquial Gujarati, the possessor is the expressed by a locative phrase headed by *pas-e* 'near-LOC', which is modified by *tem=ni* 'of him'. The possessed argument is expressed by the morphologically unflagged NP "property worth millions of rupees", which is also indexed on the verb. In (5.20b), from Palula, the possessor is expressed by the locative phrase *aɣíi=wee* 'in him', and the possessed is expressed by the morphologically unflagged NP "knife", which is also indexed on the

verb in gender and number. Similar examples are found in Hindi, where the possessor is flagged by the Genitive marker and *pas* ‘near’ or by the locative clitic *=mẽ* (see Montaut 2012:149-150 for details).

(5.20a) *tem=ni pas-e lakh-o rupiya-n-i milkət hə-t-i*
3SG=GEN.FSG near-LOC million-PL rupee-GEN-FSG property be-PST-FSG
‘he had property worth millions of rupees’ (Gujarati, Christian 1987:241)

(5.20b) *kaʈéeri bi aɽíi=wee hín-i*
knife TOP 3sg.OBL=in be.PRS-F
‘he had a knife’ (Palula, Liljegren & Haider 2015:117)

5.2.2.4 Possessor as Companion

In the data there are several examples of predicative possession with “be/become” in which the possessor is flagged by a comitative marker. This construction type is similar to Heine’s 1997 “companion schema” which he schematically described as “X is with Y”. Most of Heine’s examples, however, have the possessed argument flagged by the comitative marker. Instances in which the possessor is flagged by a comitative marker are treated there as “minor” (Heine 1997:56-57). This construction type is infrequent in the data analyzed for this study and is demonstrated below by examples from Palula and Nagamese. Other languages in which this construction type is found include Gorani and Bactrian. In example (5.21a), the possessor is flagged by the postposition *díi* ‘with’, which governs a NP in the accusative case. The possessed is expressed by the morphologically unflagged NP “two, three crates”, which is also indexed on the verb. In

(5.21b), the possessor is followed by the post-position *logote* which is often used to flag comitatives and causes.

(5.21a) *dúu tróo kareeɽ-í haan-séen-i tas díi*
two three crate-PL be-PRS-F 3SG.ACC with
'he has two, three baskets' (Palula, Liljegren & Haier 2015:126)

(5.21b) *tae logote dhan.maria.pathor ase*
3SG with winnowing.fork be.PRS
'he has a winnowing fork' (Nagamese, Luke 3:17)

5.2.2.5 Other "Be/Become" Predicative Possession Construction Types

In some of the languages described here, there are instances of clauses expressing predicative possession with "be/become" as the main lexical verb which do not exactly fit in one of the other construction types described here. That is, the synchronic grammar of these clauses differs from what is implied by the schemata proposed in Heine 1997. This is mostly due to two crosslinguistically common processes of change: "transitivization" or "have drift" and "adjectivization" (both terms due to Stassen 2009, 2013b). In this section, I briefly analyze instances where these processes have an affect on the coding properties of the clause.

5.2.2.5.1 Transitivization, "Have Drift"

In some languages, certain clause constructions expressing predicative possession with "be/become" start to slowly "gain" more and more grammatical properties associated with transitive clauses (hence, they drift towards "have" types predicative

possession). This process was briefly demonstrated in Stassen 2013b by examples from Luganda (Bantu) and Luiseño (Uto-Aztecan).

In Luganda, where predicative possession is expressed by a construction belonging to Heine's comitative schema, the comitative marker cliticizes to the "be/become" verb, and the compound word composed of these two behaves like a regular, albeit compound, transitive verb. In Luiseño the possessor is originally coded as a Topic, but can also be optionally coded as an A argument. Further, in Modern Hebrew (Semitic), predicative possession is expressed by a verbless construction classifiable as Heine's Goal Schema ("X exists to Y"), where the possessor is flagged by the same preposition that flags recipients and goals of motion. The possessed NP in this construction, however, is flagged as a P argument in the DOM system of Modern Hebrew.

All these possessive constructions gain grammatical properties associated with transitive clause constructions, and hence "drift" towards more transitive structures. Often, synchronic constructions which are in the process of such a drift cannot be easily classified as one of the major typological types of predicative possession.

In the languages analyzed here, similar processes are found which involve changes in the grammatical properties of possessors that can be interpreted as a shift or a drift towards grammatical coding associated with A arguments. These include changes in the behavioral tendencies of possessors and in their coding properties. It seems that changes in behavioral properties are more common in the data than changes in morphological coding properties of possessors. This replicates the "behavior-before-coding" principle argued for in Cole et al. 1980.

Changes in the behavioral tendencies of possessor arguments can be illustrated by the discourse status of possessors and in their mention types (lexical NPs, clitic pronouns, identifiable zeroes) in Middle Persian. As seen above in subsection 5.2.2.2, Middle Persian possessors are expressed in three different ways: by unflagged NPs, NPs flagged by *rāy* 'to, for', and by second position clitic pronouns. The first two strategies are infrequent, and the third strategy, the clitic pronoun, accounts for over 90% of the instances of predicative possession in the Middle Persian data analyzed here.

The high frequency of the clitic pronoun in possessor function can be argued to be the result of the discourse status of most possessors. For the most part, discourse revolves around the referents of possessor arguments, not possessed arguments. Thus, possessors, like A arguments, are more likely than possessed to be accessible and given in discourse (i.e., discourse topics in the sense of Givón 2001). Hence, they are more likely to be expressed by mention types associated with discourse accessible referents, such as pronominal and zero mentions.

The high frequency of clitic pronouns, as opposed to contextually recoverable zeroes, expressing possessors can be given a diachronic explanation (i.e., it has little to do directly with their discourse status as discourse topics). Middle Persian clitic pronoun are cognates of the Old Persian Genitive-Dative clitic pronoun. The Old Persian clitic pronoun was used to express possessors alongside a Genitive-Dative form of lexical nouns (see examples (5.9b) and (5.11a) above). Thus, the high frequency of possessors expressed by clitic pronouns in Middle Persian can be explained as a retention of an older pattern: the most common form used to express possessors in Middle Persian is the only surviving cognate of the forms used to express possessors in previous stages. Further

evidence for this scenario come from two very similar situations identified with verbs encoding "need" and "seem" semantics, in which the clitic pronoun is significantly more frequent than any other means for expressing the "needer" and the "seemer" microroles, even if less overwhelmingly so than it is for predicative possession.[45]

These two explanations, i.e., the higher discourse accessibility of possessors and the origin of the clitic pronoun in the old Dative-Genitive clitic pronoun, conspire to the same result: the high frequency of possessors being expressed by the clitic pronoun. The similar discourse properties of possessors and A arguments in Middle Persian, however, do not lead to a change in the coding properties of possessors to be more similar to the coding properties of A arguments. The increased similarity is limited to behavioral and discourse properties.

The reminder of this subsection argues that in Sivandi and Gorani, two Central Iranian languages, a shift occurs in the coding properties of possessors with "be/become" which ends up with possessors coded as A arguments. In these two languages, possessors in predicative possession constructions with "be/become" are often expressed in the adnominal modifier construction type demonstrated above in section 5.2.2.1. In Sivandi and Gorani, possessors are expressed by the clitic pronoun, which also functions as a possessive pronoun, attached to its functional head, as illustrated in (5.22a-b), repeated from (5.10b-c) above.

[45] See Shirtz (in preparation). This explanation can be construed as adding to Bickel 2003, who argued that roles coded by unflagged NPs are more frequently expressed by zero mentions than roles coded by flagged NPs.

(5.22a) *usūr-ā=mā bi*

horse-PL=1PL be.PST.3SG

'we have horses' (Sivandi, Lecoq 1979:152)

(5.22b) *yak kaškūł=iš h-ē*

one bag=3SG be.PRS-3SG

'he has a bag' (Gorani, Mahmoudveysi et al. 2012, 4:32)

In both Gorani and Sivandi, a clitic pronoun is used to express adnominal possessors (see Mahmoudveysi et al. 2012:16-17 for Gorani, and an abundance of examples such as *kākā=š* "his brother" in the Sivandi texts in Lecoq 1979). In addition to this function, however, in both languages the clitic pronoun is also used to index A arguments of verbs in the past tense (see subsection 1.3 above). In this function, the constraints determining the position of the clitic pronoun differ slightly between the two languages, but in the vast majority of past tense clauses in the data from both Sivandi and Gorani, clitic pronouns indexing the A arguments are preverbal, i.e., they attach to the last constituent before the verb. This is demonstrated by example (5.23) from Gorani, where the A argument *tu* is also indexed by the preverbal clitic pronoun *=it*.

(5.23) *tu min=it dī*

2SG 1SG=2SG see.PST

'You saw me' (Gorani, Mahmoudveysi et al. 2012:39, glosses mine)

When the Gorani and Sivandi clitic pronoun is used to express possessors and index A arguments it tends to be preverbal; that is, it is always preverbal in the predicative possession constructions, and mostly preverbal in the past tense construction.

If possessors in the predicative possession in both Sivandi and Gorani are reanalyzed (at least by some speakers) as A arguments, a potential actualization is that the clitic pronouns in predicative possessions become used to index possessors, rather than to refer to them as adnominal modifiers. If such a process of reanalysis-and-actualization in fact occurred, possessors would be expressible lexically in the predicative possession construction by an unflagged NP alongside the clitic pronoun.

Such examples are rare, but are attested in the data as shown in (5.24a-b). In (5.24a), from Sivandi, the Direct case form of the 1SG pronoun, *me* 'I', expresses the possessor, whereas the preverbal clitic pronoun can be said to index the possessor. In (5.24b) from Gorani, the possessor is expressed lexically by the complex NP "our mother" and is indexed by the preverbal clitic pronoun.

(5.24a) *me ye abāī=m bi*
1SG INDEF coat=1SG be.PST.3SG
'(then, he replied) I had a coat' (Sivandi, Lecoq 1979:161)

(5.24b) *na, dāyka=y ēma šāx=iš hē*
NEG mother=LNK 1PL horn=3SG be.PRS.3SG
'No, our mother has horns' (Gorani, Mahmoudveysi et al. 2012, 2:42)

Interestingly, there is a slight difference between the type of evidence for this reanalysis found in Sivandi versus in Gorani. All examples which exhibit this actualization in Sivandi are in the past tense. This is to be expected, as the clitic pronoun function of indexing the A argument is limited to the past tense. In Gorani, however, examples for the actualization of this reanalysis are available in all tenses. One possible

explanation for this situation is that initially, the actualization started with the past tense, and then could potentially be extended to other tenses. More evidence, however, is needed to positively argue for this interpretation.

Finally, it should be noted that the reanalysis of possessors as A arguments is not automatic or obligatory. As in Gorani and Sivandi, in Taleshi, another Central Iranian language, the clitic pronoun is attested in both functions of expressing adnominal modifiers as well as indexing A arguments when the verb is in the past tense (see Paul 2011). Predicative possession in Taleshi is expressed by a similar "be/become" construction, with the possessor as an adnominal modifier, as shown in (5.25a). In Paul 2011 there are no examples indicating a change in the coding properties of possessors towards that of A arguments. When possessors cannot be expressed by a pronominal mention for discourse reasons, for example, they are expressed by a post-positional phrase with *=ra*, similarly to the Middle Persian construction type demonstrated in (5.14) above.

(5.25a) *ila* *kəla=m* *hest=a*
INDEF girl=1SG exist=3SG
'I have a girl' (Taleshi, Paul 2011:253)

(5.25b) *mə=ra* *diar* *ka=ni* *hest=a*
1SG.OBL=for another house=also exist=3SG
'I have another house' (Taleshi, Paul 2011:223)

To summarize this section, in many of the languages in the sample, possessors with "be/become" share pragmatic and discourse properties associated with A arguments:

they tend to be overwhelmingly animate and human, and tend to refer to highly identifiable discourse participants. In some languages, these facts, together with the diachronic source of predicative possession constructions, lead to a situation in which possessors are expressed by the Iranian clitic pronoun significantly more often than by lexical forms. In Gorani and Sivandi, the functional range of this clitic pronoun has made possible its reanalysis as an index marker of A arguments, which in turn has led to a reanalysis of possessors with "be/become" as A arguments.

5.2.2.5.2 "Adjectivization"

Another process identified by Stassen (2009, 2013b) is what he calls "adjectivization", where the possessed NP is reanalyzed as the predicate; this reanalysis can be actualized in several different ways (see Stassen 2009, 2013b for more details). One of the cases of adjectivization cited in Stassen 2013b is from Tiwi (Australia, Isolate) where predicative possession is expressed by two morphologically unflagged NPs.

Examples of similar construction types are found in the data analyzed here. In (5.26a-b), repeated from (5.7a-b) above, the possessor and the possessed arguments are both expressed by unflagged NPs, followed by a verbal copula. In terms of coding properties these clauses are indistinguishable from clauses expressing core nominal predication. Further, as shown in (5.8) above, similar verbless clauses were also found in the data analyzed for this dissertation.

(5.26a) *harw kas ciš=ē ast*
every person thing=INDEF be.PRS.3SG
ī=š az abārīg ciš āzarmīg-tar
LNK=3SG from other thing precious-CMP
'Every person has one thing which is dearer than other things' (Middle Persian, DK6 33)

(5.26b) *ukʰ=rə dulhi=k cʰawa nidz=bʰəi-lə*
3SG=GEN wife=3SG.POSS son NEG=be-PST
'his wife did not have a son (lit. his wife was not a son)' (Darai, Dhakal 2013:79)

This type of example is not frequent in the data, either in terms of the number of languages in which it has been identified nor in terms of its frequency. The fact, however, that there are several examples from five different languages points to the possibility that processes which lead to this pattern are crosslinguistically more common than perhaps has been assumed.

5.2.2.6 Summary

Subsection 5.2.2 described and analyzed the different construction types expressing predicative possession with "be/become" as a main verbal predicate in Indo-Iranian. It has been shown that the expression of predicative possession with "be/become" in Indo-Iranian comprises a typologically diverse field, and includes several of the different schemata identified in the typological literature (following Heine 1997, Stassen 2005, 2013c); the genitive schema ("X's Y exists), the goal schema ("Y exists to / for X"), the locative schema ("Y exists in X"), and the companion schema ("Y exists

with X"). Subsection 5.2.2.5 also presented evidence for the two general processes of "have drift" and "adjectivization".

Even a cursory glance at the examples cited above would confirm that intralinguistic variation in the expression of predicative possession is the norm, rather than an exception. In many languages, several constructions with "be/become" express predicative possession, often with different schemata. The intralinguistic variation is even broader considering the fact that in several Iranian languages, verbs originally denoting the meaning of "hold" have grammaticalized and are used as "have" type verbs, either as an internal innovation, or through borrowing from Modern Persian.

As a complete analysis of the different diachronic processes behind this typologically diverse field is well beyond the scope of this section, for now I will just suggest a hypothesis of what these processes might have looked like. As shown in subsection 5.2.2.1, in the older Indo-Iranian languages, Vedic Sanskrit and Old Persian, the possessed argument was expressed by a Nominative NP and was indexed on the verb, while the possessor argument was expressed by a Genitive NP in Sanskrit and a Genitive-Dative NP in Old Persian. Following the syncretism of the Iranian Genitive and the Dative cases, the grammatical status of the possessor NP was ripe for reanalysis as an independent, external, NP constituent rather than an embedded adnominal modifier, because the same dative-genitive case forms was often associated with functions other than adnominal modification. This led to a change in the typological schema of predicative possession from Heine 1997's Genitive Schema to Heine 1997's Goal Schema, which can be seen, for example, when the possessor is expressed by the second position clitic pronoun as in example (5.11a) above. The eventual loss of the genitive-

dative case form made room for already existing, but perhaps quite marginal and infrequent, innovative adpositional markers to take over, although in Iranian the clitic pronoun was, and still is, very frequently used for expressing possessors. The typological variation in predicative possession schemata identifiable in Indo-Iranian is, then, a result of the different innovations which took over the expression of possessors at this stage along side the retention of the use of the clitic pronoun in Iranian.

5.2.3 "Be/Become" Expressing the Predicate Locative and the Existential

This subsection analyzes construction types expressing either the predicate locative or the existential meaning, in which "be/become" is the main verbal predicate. The definitions used here for these two functional domains are based on Payne 1997:111-113 and Creissels 2013, 2014. Under these definitions, a clause expresses the predicate locative functional domain if it predicates the location of some object, and a clause encodes the existential functional domain if it expresses the existence of some object in some, usually identifiable and given, location. Following Creissels 2014, I recognize that under this definition, an existential is basically an "inverse locative".

The different construction types identified are presented here in two sets. In the first, including only existential construction types, "be/become" is accompanied by a single NP, the figure. The ground is not overtly mentioned and can be either interpreted contextually, or refers to "the world". The second set of construction types includes those construction types in which "be/become" is accompanied by two NPs, one expressing the figure and the other expressing the ground.

5.2.3.1 "Be/Become" Accompanied by a Single NP

In the construction types described and analyzed in this subsection, "be/become" is accompanied by a single, unflagged NP which expresses the figure argument. In the data, the semantics of the figure NP is not limited to beings or objects with a physical reality, but the figure NP is also found encoding events and temporal (deictic) points. The semantics of the entire clause and its discourse function depend to a large degree on the semantics of the figure NP.

The most frequent type of figure NP found in this construction type expresses an animate entity or a tangible object. The clause itself expresses the existence of the figure NP either generally ("in the world") or in some identifiable, given location that is left unexpressed. This construction type is found in all languages in the sample, and is illustrated in (5.27a-c) below. In example (5.27a), from Darai, the figure is expressed by the unflagged NP *bʰat* 'cooked rice'. The ground relative to which the existence of the cooked rice is predicated is the house of those people who are the main protagonists in the story. In (5.27b), the existential clause is a complement clause reporting what one of the protagonists saw. The ground in (5.27b), then, is the here and now of the narrative. In (5.27c), the speaker describes the preparations for a celebration in the village. This celebration is the ground element, not overtly expressed, in the two existential clauses in (5.27c).

(5.27a) *bʰat bʰəi-te sag nidze*
rice be-SIM curry NEG
'If there was rice, there wasn't any curry' (Darai, Dhakal 2013:13)

(5.27b) *ta inɳʈʂ hín-u*

DS bear be.PRS-MSG

'(he looked and) there was a bear' (Palula, Liljegren & Haider 2015:96)

(5.27c) *hezār=tā e de hezār=tā e*

thousand=CLF be.PRS.3SG two thousand=CLF be.PRS.3SG

'(If) there were on thousand, two thousand (people, we cooked for everyone)'

(Ziyarat Mazandarani, Shokri et al. 2013:84)

In many languages in the sample analyzed here, one of the main discourse functions of this construction type is to introduce new participants to the discourse. This can be seen in (5.27b) above, where the bear, one of the main characters in a narrative that describes an encounter with a bear, is first mentioned. This function of this existential construction type is further illustrated in examples (5.28a-b). In both examples, the ground is not the "here-and-now" of the narrative as in (5.27b) above, but rather the general world. Example (5.28a) is the first clause of the narrative text, and example (5.28b) is the first clause in a direct speech quotation in which a king describes a pond he remembers.

(5.28a) *ye šāh-i bi, se kor=āš bi.*

one king-INDEF be.PST.3SG three son=3SG be.PST.3SG

'there was a king, he had three sons' (Sivandi, Lecoq 1979:107)

(5.28b) *goʈek bond ac-e*

INDEF pond be-PRS.3SG

'there once was a pond' (Kotia Oriya, Gustafsson 1973a:187)

In examples (5.29a-c) below, the figure NP expresses an event or a state. With the cognate "be/become" verbs analyzed here, this construction type is relatively infrequent. The discourse function of clauses in this construction type is often to provide background, or set up the necessary starting point for a stretch of discourse. Examples (5.29a, 5.29c) set up background for further discourse, discussing what people do in the winter time or what happens when there are robberies. In (5.29b), on the other hand, the ground is the specific temporal point in which the marriage of the speaker took place.

(5.29a) *sard-i* *ā-bu* *ke* *garmine* *bo-xor-i* *dige.*
cold-ABST PRV-be.PST COMP warming SUBJ-eat.PRS-2SG PRT
'If it gets cold, you should eat warming things, you know' (Ziyarat Mazandarani, Shokri et al. 2013:117)

(5.29b) *nahi-hĩ* *biha* *bʰəi-te*
young-FOC marriage be-SIM
'as my (young) marriage took place' (Darai, Dhakal 2013:106)

(5.29c) *ɖaakeé* *ta* *bh-íl-a* *seentá*
robberies DS be-PERF-MPL COND
'at the time of robbery (they took our guns)' (Palula, Liljegren & Haider 2015:22)

In examples (5.30a-c) below, the figure NP refers to a temporal point. This is a much more common construction type in the data, found in most of the languages analyzed here. This construction type expresses the passing of discourse time until the point in time expressed by the figure NP, or that the time interval expressed by the figure NP has passed. The first type of clause is illustrated by (5.30a-b), where the figure NPs

are “early dawn” and “midnight,” respectively. Example (5.30c) demonstrates the second type of clause where the figure NP is “one month”.

(5.30a) *vaqt=i ke rassā nazīk=i ū kākā=š,*
time=LNK REL arrive.PST close=LNK DEM brother=3SG
dame šafaq bi
early dawn be.PST.3SG
‘when he got close to his brother, it was early dawn’ (Sivandi, Lecoq 1979:89)

(5.30b) *baaro gunta rati oi-la*
twelve hour night be-PST.3SG
‘it became midnight’ (Kotia Oriya, Gustafsson 1973a:122)

(5.30c) *tā ē māh be baw-ēd*
until one month PRV be.PRS-3SG
‘until one month has passed’ (Middle Persian, PRDD 2)

5.2.3.2 “Be/Become” Existential with a Relative Clause

In this construction type, “be/become” is followed, rather then preceded, by an optional lexical NP and an obligatory relative (nominalized) clause. This relatively infrequent construction type is illustrated in examples (5.31a-b) from Middle Persian and Sogdian, where its function is usually to introduce a new character to the discourse and characterizes it through the content of the obligatory relative clause.[46] In (5.31a), the figure NP *dastwar* ‘sage’ is followed by a relative clause describing the opinion of that

[46] This is quite similar to the function of the American English construction illustrated in *there was a farmer had a dog* as analyzed by Lambrecht 1988.

sage in some religious matter. In (5.31b), *ast* is followed by a relative clause headed by the Ezafe Linker *ī*. It described the manner in which people cross a mythical river in the afterlife. In (5.31c), from Sogdian, *bʔt* 'be' is followed by the NP "animate being" which is modified by a relative clause.

(5.31a) *būd dastwar kē=š guft kū ...*
be.PST.3SG sage REL=3SG say.PST.3SG COMP …
'There was a sage who said: …' (Middle Persian, PRDD 15a)

(5.31b) *ud ast ī pad garān ranj hamē widard*
and be.PRS.3SG LNK by grave pain PRV cross.PST
ud ast ī xwārīhā widard h-ēnd
and be.PRS.3 LNK easily cross.PST be.PRS.3PL
'and some (people) were crossing it with great difficulty and some (people) were crossing it easily' (Middle Persian, AWN 16.2)

(5.31c) *rty kð bʔt wʔtðʔr ʔʔʒʔwn ʔky ...*
and when be.PRS.SUBJ.3SG animate being REL …
'and if there is an animate being which…' (Sogdian, MacKenzie 1976. DHY 235)

5.2.3.3 "Be/Become" with a Locative Phrase and an Unflagged NP

This subsection describes two construction types in which "be/become" is accompanied by an unflagged NP encoding the figure and another NP flagged by some locative marker encoding the ground. These construction types were found in all sixteen languages analyzed here.

This configuration of coding means most commonly expresses the predicate location and the existential functional domains. In Palula and Nagamese this

configuration also expresses predicative possession, and at times, this construction type is also extended to express the core nominal predication, usually together with a metaphoric extension of the locative ground to express a state (e.g., *she is in happiness)*, or of the figure NP to express a state and the ground NP to express an entity (e.g., *happiness is in her*).

The use of this configuration of coding means to express the existential function is illustrated by (5.32a-c). In (5.32a), from Pali, the ground is expressed by "below the rock" and the figure by "water" in the Instrumental case.[47] In (5.32b), from Kotia Oriya, the ground is flagged by the Locative marker *–e,* and in (5.32c), from Sivandi, the ground is flagged by the preposition *dar* 'in, inside'.

(5.32a) *pāsāṇassa heṭṭhā udakena bhavitabban*
rock.GEN below water.INST be.FUT.PASS.PTCP
'there must be water under that rock' (Pali)

(5.32b) *se bitr-e goʈek boti ac-e*
DEM inside-LOC INDEF lamp be-PRS.3SG
'inside it (the temple) there's a lamp' (Kotia Oriya, Gustafsson 1973a:127)

(5.32c) *dar ešk̆aft=i nūr=i čerāγ=i en*
in cave=INDEF light=LNK lamp=INDEF be.PRS.3SG
'(he saw that) in the cave, there was the light of a lamp' (Sivandi, Lecoq 1979:108)

[47] The Instrumental case is required for S/A arguments when the main verbal predicate is in one of the passive participle forms. Note, however, that the term "passive" is based on the historical source of the verb form and not on its synchronic function (Peterson 1998:92-99).

The same flagging and indexing pattern also expresses the predicate locative, as shown in (5.33a-c). In (5.33a), from Bactrian, the locative ground is the city, or kingdom, of Gandhara, and is flagged by the preposition *ab* 'in', which often cliticizes to the preceding NP. The figure participant is expressed by the unflagged NP *naʃalm* 'throne'. In (5.33b), from Kotia Oriya, the locative ground is expressed by "this land" and is flagged by the Locative marker *-e*. In (5.33c), the locative ground is expressed by the prepositional phrase *tu asām* 'in my pocket'.

(5.33a) *od naʃalm mal=ab ganda{ro}r=ind*
and throne here=in gandhara=be.PRS.3PL
'and (his) throne is here in Gandhara' (Bactrian, Sims-Williams 2001, W2')

(5.33b) *pila se des-e ac-e*
body DEM land-LOC be-PRS.3SG
'the boy was in that land' (Kotia Oriya, Gustafsson 1973a:199-200)

(5.33c) *pīl=em tū asā=m en*
money=1SG in pocket=1SG be.PRS.3SG
'my money is in my pocket' (Sivandi, Lecoq 1979:159)

Examples (5.32 – 5.33) differ both the nominal predication functional domain they express (the existential vs. the predicate locative) and in the relative word order of the figure and the ground. In (5.32a-c), expressing the existential function, the relative word order is ground – figure, and in (5.33a-c) that expresses the predicate locative function the relative word order is figure – ground. The correlation between these

functions and the difference in relative word order holds quite well throughout the data, which replicates the findings reported in Clark 1978.[48]

There are, however, some instances which violate this correlation; their effect on the co-expression patterns of the predicate locative and the existential functions is assessed in Chapter VIII. In (5.34a-b) from Kupia, the relative word order is the same, ground – figure – verb, but the clause in example (5.34a) predicates the location of the princess, and the clause example (5.34b) predicates the existence of a tiger in the area where the speaker lives.

(5.34a) *kaḍayi=ci antastumu=te ja na:ḍ-i as-e*
uppermost=GEN story=LOC DEM child-F be-PRS.3SG
'that girl is on the top story, (that princess)' (Kupia, Christmas & Christmas 1973a:80)

(5.34b) *palana=te wa:gu as-e*
area=LOC tiger be-PRS.3SG
'there's a tiger in the area' (Kupia, Christmas & Christmas 1973a:42)

In the vast majority of the examples found in the data, as seen in examples (5.32) – (5.34), the ground phrase is either spatial or temporal. There are, however, instances of this configuration of coding means in which the locative phrase is used to express a state or a property, and thus, the clause does not express the existential or the predicate locative functional domains. Instead, it expresses the proper inclusion or the predicative

[48] Clark 1978 motivates this difference by information flow considerations. Given that all the languages analyzed here are verb-final and thus might prefer to have the "given" information earlier in the clause, the word order difference identified in the data here should not be entirely surprising.

property functional domains.[49] In (5.35a), the figure (referring to a dragon) is not overtly expressed in the clause. The ground, however, is overtly expressed and is flagged by the post-position *madʒí* 'in'. In (5.35b), the figure NP is "tailoring" and the ground phrase is flagged by the preposition *bā* 'in'.

(5.35a) *níndram madʒí hín-u*
sleep in be.PRS-MSG
'(The dragon) was in sleep' (Palula, Liljegren & Haider 2015:78)

(5.35b) *xayyāt-i bā in barnāme be*
tailor-ABST in DEM plan be.PST.3SG
'Tailoring was like this' (Ziyarat Mazandarani, Shokri et al. 2013:149)

Alongside the metaphoric extension of locations to states, there are instances in which the figure NP encodes a state, and the locative phrase expresses the person, or element, that is in that state. This is demonstrated in (5.36a-b), from Sivandi and Palula.

(5.36a) *ī ye serr-i dar kār-e h-and*
DEM INDEF secret-INDEF in deed-DEF be.PRS-3SG
'there's an enigma in this affair' (Sivandi, Lecoq 1979:128)

(5.36b) *eetieé ma dʒhulí asár bhíl-i*
then 1SG on effect be.PRF-FSG
'then, I was afflicted (lit. the effect was on me)' (Palula, Liljegren & Haider 2015:46)

[49] This is one of the rare instances in which core nominal predication domains differ in the data. There are no instances in the data where the locative / existential construction type expresses equation.

Examples for these two metaphoric extensions of the configuration of coding means usually associated with the predicate locative or the existential functional domains are not uncommon with the cognate "be/become" verbs analyzed here. The affect that these extensions have on the overall co-expression patterns of nominal predication functions is analyzed in Chapter VIII.

5.2.4 Complex Predicate Usages

There are several instances in the data where "be/become" is used as the main verbal predicate in clauses not expressing any of the six nominal predicate functional domains. In these instances, "be/become" is accompanied by three elements instead of one or two, and the lexical predicative concept expressed in the clause is usually encoded by a pre-verbal element, similar to the deed element found in many "do/make" construction types described in Chapter IV. This preverbal element encodes usually modal or knowledge semantics.

While this construction type is illustrated by three examples from Sivandi, Ziyarat Mazandarani, and Gorani, all Central Iranian languages, it is also attested in some Indo-Aryan languages including Hindi and Nagamese. In (5.37a-b), from Ziyarat Mazandarani, "be/become" is preceded by *motevaje* 'aware, realizing' (a Semitic loanword). In the glossary published by Shokri et al. (2013), this word is classified as an adjective,[50] with a subentry for *motevaje* plus "be/become" meaning "to understand". In (5.37a-b) "be/become" is also accompanied by a 2SG pronoun flagged by DOM, expressing the

[50] All examples for *motevaje* in the data analyzed and supplied in Shokri et al. 2013 have the word used in this complex construction. There are no examples of *motevaje* functioning as an adnominal modifier.

person who understands something. The contents of the knowledge are expressed in (5.37a) by a finite complement clause which follows "be/become", and in (5.37b) by the NP *kārem* 'my deed'.

(5.37a) *yani šervār alān ma te=re masan motevaje ā-be*
that.is trousers now 1SG 2SG=OBJ for.example aware PRV-be.PRS.3SG
čejur me-rviniy-em dige.
how IMPF-cut-1SG PRT
'that is, the trousers, I (want) you to know how I cut them' (Ziyarat Mazandarani, Shokri et al. 2013:153)

(5.37b) *te=re motevaje kār=em bu*
2SG=OBJ aware deed=1SG be.PST.3G
'I will make you aware of my work (lit. you were aware of my work)' (Ziyarat Mazandarani, Shokri et al. 2013:152)

In examples (5.38a-b), from Sivandi, "be/become" is preceded by *balad* 'known' (another Semitic loanword). Apart from *balad*, "be/become" is accompanied by two elements. The first expresses the knower, which as shown in (5.38a-b), is also indexed on the verb. The second element expresses the known contents, which are expressed in (5.38a) by a complement clause headed by *ke*, similar to Object Complement clauses in Sivandi, and in (5.38b) by the NP "his house", which is flagged by DOM.

(5.38a) *fey-em balad ām ke če felez-xarabi en*
REFL-1SG known be.PRS.1SG COMP what grain-bad be.PRS.3SG
'I know what bad grain this is' (Sivandi, Lecoq 1979:88)

(5.38b) *diye=šā=rā balad i?*
house=3SG=OBJ known be.PRS.2G
'do you know his house?' (Sivandi, Lecoq 1979:138)

Finally, in (5.39a-b) from Gorani, "be/become" is preceded by *garak* 'necessary', a Turkic loanword. Apart from *garak*, "be/become" is accompanied by two elements. In (5.39a), the needer is expressed by a clitic pronoun and the needed is expressed by NP. In (5.39b), an elicited example from the sketch provided in Mahmoudveysi et al. (2012:44-45), the needed element is expressed by a complement clause with a subjunctive main verb.

(5.39a) *pīknīkēk=u hüčtān=im dī garak nīya*
gas.cooker-INDEF=and nothing=1SG then necessary be.PRS.NEG.3SG
'(bring) a small gas cooker, and I need nothing else (from you)' (Gorani, Mahmoudveysi et al. 2012, 3:86)

(5.39b) *min garak=m=ē kar bi-kar-im*
1SG necessary=1SG=be.PRS.3SG deed SUBJ-do.prs-1SG
'I want to / I must work' (Gorani, Mahmoudveysi et al. 2012:45)

5.3 "Be/Become" as Non-Main Verbal Predicate

In all the construction types described so far in this chapter, "be/become" is the main, and only, formal verbal predicate of the clause. This includes the construction types used to express the six nominal predication functions and the construction types in which "be/become" verbs form a complex predicate together with another lexical element. In

contrast to the preceding sections, this section sets forth the construction types in which "be/become" functions as an auxiliary verb in a periphrastic verbal construction. For the most part, these constructions are used to express different TAM and voice functions. The major criterion used here to distinguish between construction types is the grammatical function they encode. We will see that "be/become" is used in auxiliary constructions encoding tense (past, present, and future tense), perfect aspect (or Aktionsart), subjunctive and obligative mood, and passive (or passive-like, P-oriented) voice. For the most part, these construction types have been well described in grammars, grammatical sketches, and by dedicated papers, and this chapter is mostly based on these descriptions.

This sketch is, perhaps somewhat artificially, limited in two ways. First, it abstracts over some language-specific and construction-specific structural differences. These differences have to do with the form of the main lexical verb (finite verb form? infinitive? participle? gerund?), and with the relationship between the main lexical verb and the "be/become" auxiliary. For example, in Middle Persian, a single "be/become" auxiliary verb in the past tense auxiliary construction can have scope over two or, rarely, three clauses, each with a different main lexical verb. In other periphrastic constructions across the data, the cohesion between the main lexical verb and the auxiliary verbs is tighter, with each auxiliary having scope over only one main lexical verb.

The second way in which the sketch here is limited is that the constructions included are those in which "be/become" cognate verbs are still identifiable as verb words, and have not been affixed to the main lexical verb in any way. Thus, this section does not include instances where "be/become" auxiliary verbs have undergone further

phonological reduction to become affixes, resulting is a synchronically simple verb form. This occurred in daughter (or niece) constructions of the Middle Persian past-tense auxiliary construction described below. In Early Judeo-Persian and other varieties of New Persian, the "be/become" auxiliary construction underwent extreme reduction resulting in a suffixation of the old "be/become" auxiliary to the main lexical verb. This is found, at least in all varieties of New Persian, including Early Judeo-Persian (Gindin 2008, Paul 2013), Classical Persian (Lazard 1963), and Standard New Persian.

5.3.1 Periphrastic Tense Expressing Constructions

Periphrastic constructions with "be/become" in the data often express present, past, or future tenses. Examples (5.40a-b) below, from Hindi, demonstrate the use of "be/become" as an auxiliary verb in constructions expressing habitual present tense (Kachru 2006:82, Montaut 2012:115-116). In example (5.40a) from Montaut's 2012 grammar sketch, the "be/become" auxiliary verb is preceded by the main lexical verb, a participle form of "drink". The number and roles of arguments are determined by this main lexical verb. As Montaut 2012:116 notes, however, and as (5.40b) illustrates, this Hindi construction is also used to express events in the immediate future.

(5.40a) *ām.taur.par kāfī pītā hũ̃*
usually coffee drink.PTCP be.PRS.1SG
'I usually drink coffee' (Hindi, Montaut 2012:116)

(5.40b) *maĩ abʰī ātā hũ̃*
1SG immediately come.PTCP be.PRS.1SG
'I shall come immediately' (Hindi, Montaut 2012:116)

Auxiliary constructions expressing the future tense are less common in the data. They are illustrated here by examples from Nagamese, where the present form of "be/become" is preceded by the *–bole* participle of the main lexical verb. The number, coding properties and semantic roles of participants are determined by the main lexical verb. In (5.41a), *ase* 'be/become' is preceded by *dibole*, the *-bole* participle form of 'give', which motivates the ditransitive grammar of the entire clause.[51] In example (5.41b), the main lexical verb is "laugh" and the clause has intransitive grammar. This form is not mentioned in Sreedhar 1985.

(5.41a) *aru apuni=laga maiki Elizabeth ekta chokora=ke jonom di-bole as-e*
and 2SG=GEN woman Elizabeth one son=OBJ birth give-PTCP be-PRS
'Your wife, Elizabeth, will give birth to a boy' (Nagamese, Luke 1:)

(5.41b) *tae-khan hasi-bole as-e*
3-PL laugh-PTCP be-PRS
'they will laugh' (Nagamese, Luke 6:21)

In several languages in the data, auxiliary constructions with "be/become" verbs also express past tense, as illustrated by examples from Middle Persian. This Middle Persian construction, it should be noted, is the main locus of ergative alignment (or perhaps better, non-nominative alignment) in Middle Persian. In examples (5.42a-b), the auxiliary verb is *hēm*, the 1SG form of "be/become". In (5.42a), the clause expresses a motion event and is an intransitive clause. The auxiliary verb indexes the S argument, the

[51] The combination "give birth" in (5.41a) is common in the Nagamese data, and could very well be an English calque since English is the official language of Nagaland, the Indian state where Nagamese is spoken.

moving figure. In example (5.42b), the Middle Persian Ergative construction, the auxiliary verb indexes the P argument.

(5.42a) *ka ō ānōh frāz mad h-ēm*
when to there forth come.PST be.PRS-1SG
'when I came there' (Middle Persian, AWN 4.5)

(5.42b) *ā=t nēk-tar kard h-ēm*
then=2SG good-CMP do.PST be.PRS-1SG
'and you made me better' (Middle Persian, AWN 4.14)

5.3.2 Periphrastic Constructions Expressing Mood

In some languages, "be/become" verbs are used as auxiliaries in constructions expressing modal semantics. This use of "be/become" is less common than its use in periphrastic constructions expressing tense or aspect / Aktionsart. Here, I illustrate this use by the Past Subjunctive construction in Ziyarat Mazandarani and the Obligative construction in Hindi.

The Subjunctive mood is expressed in Ziyarat Mazandarani by two different types of constructions: a simple finite verb form, which is claimed to be a copy from New Persian (Shokri et al. 2013:39), is used to express the Non-Past Subjunctive function; an auxiliary verb construction with "be/become" expresses the Past Subjunctive function. Shokri et al. (2013:46) describe the structure of this auxiliary form as composed of the Past Participle form of the main lexical verb, followed by the Non-Past Subjunctive form of "be/become". In their analysis, "be/become" is marked as cliticized to the main lexical verb. This construction is shown in (5.43) below.

(5.43) *age ye vaxt=e ba-ʃ-e=bi*
if one time=INDEF PREF-go.PST-PTCP=be.SUBJ.PRS.2SG
'If you had gone there' (Ziyarat, Shokri et al. 2013:46)

Another example for "be/become" functioning as an auxiliary verb in clauses expressing modal semantics is the Hindi Obligative constructions (Kachru 2006:209-210, Montaut 2012:126, 145-147). In this construction, the "be/become" auxiliary is preceded by the Infinitive form of the main lexical verb. The number of participants, and their semantic roles are determined by the main lexical verb. The coding properties of the S/A argument, however, are determined by the construction as a whole, and it is flagged by the Dative marker. In (5.44) below, from Kachru's grammar, the S argument is expressed by *mujʰe*, the 1SG Dative form. Had this been simple a future tense clause, the S argument would have been in the Direct case. Montaut 2012:145-146 argues that the Dative form of the S/A arguments in this construction is related to the use of the Hindi Dative in expressing the experiencer in other Hindi constructions, such as those expressing procedural knowledge or ability.

(5.44) *mujʰe ləkʰnəū jana hoga*
1SG.DAT Lucknow go.INF be.FUT.MSG
'I will have to go to Lucknow' (Hindi, Kachru 2006:209, glosses slightly altered)

5.3.3 Periphrastic Constructions Expressing Aspect / Aktionsart

Across the Indo-Iranian language family, "be/become" verbs are often used as auxiliary verbs in periphrastic constructions expressing aspect or Aktionsart semantics. Commonly, such constructions are used to express perfect and past perfect (pluperfect)

semantics. The definitions for perfect and past perfect by different linguists working on Indo-Iranian languages seem to correlate with each other quite closely, and the textual use of the forms labeled by these terms in grammars and grammatical sketches appears to correlate as well. The definition for perfect aspect used here follows that given in Payne 1997:239-240, who characterizes perfect aspect as describing "the relevant state brought about by the situation (normally an event) expressed by the verb".

Examples (5.45a-b) below illustrate the use of periphrastic constructions with "be/become" to express the perfect aspect in Palula and Ziyarat Mazandarani. In (5.45a), from Palula, the Perfective form of "come down" or "descend" is followed by the Present tense form of "be/become". The main and the auxiliary verbs in this construction index S, P, and T arguments in intransitive, transitive, and ditransitive clauses. In (5.45a), the main lexical verb and the auxiliary verb index the S argument. In example (5.45b), the main lexical verb is in its participle form and is followed by the "be/become" auxiliary verb which also indexes the S/A argument. This Ziyarat Mazandarani construction is used to express the Past Perfect, whereas the Present Perfect is expressed by a simple verb form and no auxiliary.

(5.45a) *míi* *baábu bi* *wháat-u* *hín-u,*
1SG.GEN father also come.down.PFV-MSG be.PRS-MSG
salaám *th-íi* *de*
greeting do-3SG PST
'My father has also come, and he was telling you his greetings' (Palula, Liljegren 2008:222)

(5.45b)*pārsāl zemestān biy-ārd=bim zirxāne=mān*
last.year winter PRF-bring.PST=be.PST.1PL basement=1PL
'Last winter we took them to our basement' (Ziyarat, Shokri et al. 2013:45)

Apart from expressing the perfect, periphrastic constructions with "be/become" auxiliary forms also express the progressive aspect, albeit less commonly. Example (5.46) from Gorani shows that in the Gorani Past Progressive construction, the "be/become" auxiliary is followed (not always directly) by the main lexical verb in the finite Past Imperfective form. Both the auxiliary and the main lexical verb index S/A arguments.

(5.46) *min bī-m ma-raft-īm*
1SG be.PST-1SG IMPF-go.PST-1SG
'I was (just) going' (Gorani, Mahmoudveysi et al. 2012:43)

5.3.4 Voice Coding Constructions and the Nagamese V-*kene Be/Become* Construction

Periphrastic constructions with "be/become" auxiliary verbs also express passive voice. Passive voice is defined here, following Comrie 1989 and Givón 1984:164, as constructions encoding the de-topicalization of the agent (or other roles which are potentially coded like agents). This is a functional definition, referring only to the semantic, or functional, pole of the construction. The syntactic, or formal, pole of constructions can differ in the treatment of participants, leading to the observed crosslinguistic variation in passive marking (see Givón 2001, Payne 1997:204-216).

This functional, rather then syntactic oriented definition is often at odds with traditional descriptions of passives in the Indo-Iranian literature. First, some ergative constructions have been, erroneously, described as passives, and do not fit this definition. Second, some constructions not usually described as passive would fit this definition.

Many Indo-Iranian languages exhibit ergative constructions which have been often described as "passive constructions" (e.g., Heston 1976 for Middle Persian).[52] The use of the term "passive" to describe these constructions seems to have been motivated by three properties. First, the main lexical verb form used in these constructions is a cognate of the Old Indo-Iranian Past Passive Participle; second, the A argument is not expressed as an unflagged NP, but by some other, sometimes historically oblique, form; third, the P argument is expressed as an unflagged NP and is also indexed on the verb. Thus the definition for passive used in these publications is structural. The Middle Persian Past Tense with "be/become" auxiliary verb fits this structural definition, but it does not fit the functional definition used here.

Just as some constructions included under the syntactic definition for passive voice are excluded under a functional definition, some constructions included under a functional definition would be excluded under the syntactic one. For example, in many Iranian languages unknown, unspecified, or otherwise unimportant agents are not expressed in the clause, and verbal indexation (if obligatory in that construction) marks 3PL. Such constructions have been identified in Ziyarat Mazandarani (Shokri et al. 2013:47), but they were already occurring in Middle Persian.

---

[52] These publications predate Dixon's 1977 and 1994 publications about ergative constructions in Australian languages.

"Be/become" as an auxiliary verb in a periphrastic construction expressing passive voice is rather uncommon in the data.[53] It has been identified, however, in several languages, as illustrated here in (5.47) from Ziyarat Mazandarani. In this example, "be/become" is cliticized to the main lexical verb "sell", and the person doing the selling is not overtly mentioned in the clause.

(5.47) *zamin-ā=i ke ba-rut=ā-be*
land-PL=INDEF REL PREF-sell.PST=PRV-be.PST.3SG
'the plots of land that were sold' (Ziyarat Mazandarani, Shokri et al. 2013:47)

Another construction whose function partially overlaps with the functional definition of the passive is the Nagamese "V-*kene* be/become" construction which expresses states and resultative states. In most instances in the data, the agent is not overtly expressed and is often unknown, as illustrated in (5.48a). In some instances, however, where the construction expresses a resultant state, where the A argument of the clause is the participant that has undergone a change of state, it is overtly encoded by an unflagged NP, as shown in (5.48b). In this example, the speaker states the cause for his happiness, which is expressed by an oblique NP. This form is not mentioned in Sreedhar 1985.

(5.48a) *probitro-kitab-scripture-te likhi-kene as=e*
holy-book-scripture-LOC write-PTCP be-PRS
'In the scripture it is written (that…)' (Nagamese, Luke 4:4)

[53] For the most part, Indo-Iranian languages which express passive voice by means of a periphrastic construction do so with a motion verb that functions as an auxiliary.

(5.48b) *apuni logote ami bisi khusi pai-kene as-e*
2SG with 1SG very happy get-PTCP be-PRS
'and because of you, I became happy' (Nagamese, Luke 3:22)

5.4 Summary

This chapter described and analyzed the different uses of cognate "be/become" verbs in the data sampled for this dissertation. It was shown that cognate "be/become" verbs are used as both main verbal predicates and as auxiliary verbs. When they are used as the main verbal predicate of the clause, they are associated with several distinct construction types. These construction types differ in the number and coding of core constituents of the clause, and in the function of the entire construction. The functions such clauses express are mostly the six nominal predication functions. In some languages, however, "be/become" verbs are also found in complex predication constructions expressing modal of knowledge based semantics. "Be/become" verbs also function as auxiliary verbs across Indo-Iranian. These auxiliary construction types encode several distinct TAM categories, and less commonly encode passive (or P-oriented) voice constructions.

CHAPTER VI

CONSTRUCTION TYPES OF "GIVE" COGNATE VERBS

This chapter analyzes the different usages of "give" cognate verbs in a sample of sixteen Indo-Iranian languages. As in the previous two chapters, the construction types in which cognate "give" verbs are used are first distinguished by the function of these cognate verbs: either as the main verbal predicate or as an auxiliary verb in a periphrastic construction. Constructions of the former type are further distinguished by the number, flagging, and function of their nominal components, while constructions of the later type are distinguished based on the function of the periphrastic construction. Unlike "do/make" or "be/become", however, cognate "give" verbs are also occasionally found in what could be described as serial verb or verb seriation constructions. This indeterminacy results from the rather messy definition for serial verbs. The serial verb construction type is considered in subsection 6.3 below.

6.1 The Etymological Source and Development of "Give"

Across Indo-Aryan, cognates of PIE *$*daH_3$* 'give' are clearly attested. In Iranian, however, the loss of PIE aspirated consonants has led to the loss of distinction between PIE *$*daH_3$* 'give' and PIE *$*d^heH_1$* 'put, place' (Pokorny 1959:233-235), both reconstructed as Proto-Iranian **daH* by Cheung (2007:43-46). The Proto-Iranian form **daH,* however, is argued by Cheung to encode several further meanings, all with potentially different PIE etymological sources. These include "divide, distribute", "suck, suckle", "bind", and "beat, strike" (which reconstruction Cheung accompanies by a

question mark). Thus, Cheung reconstructs **daH* in proto-Iranian as an instance of a rather rich homophony.

Three of the six verbs that Cheung reconstructs as Proto-Iranian **daH* are never, or at least hardly ever, documented as verbs across the Iranian languages. For the "suck, suckle" meaning of Proto-Iranian **daH*, for example, Cheung cites verbs with this meaning only from modern Eastern Iranian languages, and most of cognates he cites mean "wet-nurse, foster mother" or "cow". For the meaning "bind", he cites only an Avestan verb meaning "bind" and a number of verbs from modern Eastern Iranian languages which mean "open". In some other languages, this verb survived only as a noun with meanings such as "trap". For the meaning "distribute, divide", Cheung cites only an Avestan verb. In contrast, three of the Proto-Iranian meanings of **daH* are documented as verbs across Iranian and across the Iranian languages analyzed in this study. Two of these meanings are "give" and "put, place", each with a clearly distinct PIE root; the third is "beat, hit, strike", for which no distinct PIE root has been offered.

Thus, according to Cheung (2007), cognates of Proto-Iranian **daH* have distinct sources and can be thought of as an example of homophony resulting from historical phonological processes such as the loss of aspiration of PIE aspirated stops in Iranian. In this section I will show briefly that in at least some languages, cognates of Proto-Iranian **daH* behave like a single verb in terms of clausal morphosyntax and semantics: i.e., that in at least some languages, cognates of **daH* which have more or less clear "put, place" or "beat, hit" semantics are deployed in the same clause types as cognates of Proto-Iranian **daH* that express transfer semantics. The type of event (transfer, caused change of location, or "beat, hit") is disambiguated either by the context or by the lexical

semantics of the different participants. At least in these languages, then, there are good reasons to consider cognates of **daH* as instances of synchronic polysemy which evolved out of diachronic homophony. This point is illustrated by examples from two languages: Middle Persian and Gorani.

Two constructions with the Middle Persian cognate of Proto-Iranian **daH* have semantics associated, at least potentially, with that of "place, put." In these instances, Middle Persian *dād* is the main verbal predicate in clauses expressing creation and caused change of location. These usages are shown in (6.1a-b). In (6.1a), *dād* is in a transitive clause expressing the creation of the world (the speaker being the main Zoroastrian god). In (6.1b) it is used in a clause expressing caused change of location in which the agent, the speaker, has caused the patient or theme participant (Wind) to change its location (i.e., to become the ground). The new location of the theme argument is expressed by a prepositional phrase headed by *pad* 'in, by, at'. The Middle Persian cognate of **daH* in clauses expressing transfer is illustrated in (6.1c). There, the recipient is expressed by a PP headed by the complex preposition *be ō* 'to, for, towards'.

(6.1a) *cē ka=m gēhān dād ā=m mēzd=ēw be kard*
for when=1SG world give.PST.3SG then=1SG offering=INDF PRV do.PST.3SG
'for when I created the world, I made an offering' (Middle Persian, PRDD 16.b)

(6.1b) *ud man āxist h-om u=m pad zamīg be dād*
and 1SG rise.PST be.PRS-1SG and=1SG in earth PRV do.PST.3SG
'and I rose, and put him (Wind) on the earth' (Middle Persian, PRDD 18)

(6.1c) *ān sūr ī pid ī man ohrmazd be.ō man dād*
DEM meal LNK father LNK 1SG Ohrmazd to.to 1SG give.PST.3SG
'that meal which my father Ohrmazd gave me' (Middle Persian, PRDD 26)

In each example, the number and form of the arguments accompanying the Middle Persian cognate of ***daH* are different. In (6.1a) it is accompanied by A and P arguments; in (6.1b-c) it is accompanied by three arguments. In (6.1b), as noted above, the new location of the theme is flagged by the preposition *pad*, and in examples like (6.1c) the recipient is flagged by the complex preposition *be ō*, or, more often, by *ō* alone. Thus the relationship between these meanings of Middle Persian *dād*, creation, caused change of location, and transfer, seems a clear case of homophony. The homophony is historically motivated, as Proto-Iranian ***daH* is the daughter of two distinct Proto-Indo-Iranian, and originally Proto-Indo-European, verbs: "give" and "place, put". The homophony also seems to be synchronically motivated, as *dād* is associated with a different argument frame in each of the senses.

There is, however, far from perfect correlation between the argument frames which accompany Middle Persian *dād* and the semantics of the event expressed by the clause. Events of creation are expressed by the same argument frame as events of transfer; events of transfer are expressed by the same frame as events of caused change of location; and so on. Such examples suggest that the scribes writing Middle Persian reinterpreted *dād* as a single polyfunctional verb, rather than as two or more homophonous verbs.

Events of creation involve two participants: the creator and the created element. This means that there is no "semantic room" for a recipient argument in clauses

expressing such events. In many clauses of creation, however, especially when the creator is divine or otherwise non-human, a recipient participant is overtly expressed by a general phrase roughly translated as "to the world". This is shown in (6.2a), where the giver is cultivated land, giving fruit to the world. In events of creation by a divine creator, a general "to the world"-type recipient is also often overtly expressed, as in (6.2b). Events of creation in Middle Persian, then, are sometimes expressed by clauses with the same verb and the same argument frames that are associated with events of transfer.

(6.2a) *be warz band-ēd ud bar ō gēhān dah-ēd.*
to cultivation bind.PRS-3SG and fruit to world give.PRS-3SG
'(he) ties up (land) for cultivation and it gives fruit to the world' (Middle Persian, DK6 314)

(6.2b) *ud man dām jud az ātaxš be.ō gētīg dād nē tuwān*
and 1SG creature other from fire to.to physical.world give.PST NEG able
'and I cannot create the creatures of the material world without fire' (Middle Persian, PRDD 18)

The regular expression of recipients with *dād* in Middle Persian is by a PP headed by *ō* or *be.ō* 'to'. These prepositions are distinct from *pad* 'in, at, by', which is commonly used to express the new location in caused change of location clauses with *dād*. There are, however, instances where recipients are indeed expressed by a PP headed by *pad*, and not *ō*, together with *dād* 'give'. This is commonly found when the state or status of the recipient or theme would change as a result of the transfer event, or when the recipient is a non-prototypical recipient—but, crucially, not a physical location. In example (6.3a) below, a father gives his daughter to her husband for the purpose of

marriage. The recipient is expressed by *pad šōy* 'in, at husband'. In (6.3b), the recipient is the soul of the giver, also expressed by a PP headed by *pad*.

(6.3a) *duxt ... pidar pad šōy be dah-ēd*
daughter …father in husband PRV give.PRS-3SG
'(his own) daughter the father gives to this husband' (Middle Persian, RAF Q62)

(6.3b) *ka=š drahm=ēw pūl pad ruwān be dah-ēnd*
then=3SG drachma=INDEF money in soul PRV give.PRS-3PL
'unless they give their souls a drachma…' (Middle Persian, PRDD 65)

The preposition *pad* is also commonly used as the head of a resultative PP, expressing the status or state of the theme after or during the transfer event or the state in which the theme was created. This is shown in (6.4a-c) below. In (6.4a) the cattle is given *pad dāšn ī ahlaw.dād* 'as a righteous gift'; in (6.4b), expressing creation, the sun is created *pad rawišn* 'in motion'; and in (6.4c), the woman gives herself *pad zanīh* 'in marriage (lit. in womanhood)'.

(6.4a) *ka=š hazār mēš ī mādag kē=š warrag abāg pad dāšn*
when=3SG thousand sheep LNK female REL=3SG lamb with in gift
ī ahlaw.dād be.ō mard ī ahlaw dād h-ē
LNK merit.give.PTCP to.to mand LNK merit give.PST be.PRS-3SG.SUBJ
'when he gives a thousand ewes with their lambs as a gift to a righteous person'
(Middle Persian, PRDD 16a)

(6.4b) *sidīgar xwaršēd pad raw-išn dād*
third sun in go-NMZ give.PST.3SG
'third, he created the sun in motion' (Middle Persian, PRDD 52)

(6.4c) *zan-ēw ī purnāy ke tan pad zan-īh*
woman=INDF LNK mature REL body in woman-ABST
bē.ō mard=ēw ī purnāy dah-ēd
to man=INDEF LNK mature give.PRS-3SG
'a mature woman who gives herself in marriage to a mature man' (Middle Persian, RAF Q16)

A PP headed by *pad* can be used, then, to express three related semantic notions with *dād*: (a) new location in caused change of state clauses, as in *pad zamīg* 'to the ground' in (6.1b) above; (b) state in which a theme is created or given as in *pad rawišn* 'in motion' in (6.4b) above or *pad zanīh* 'in marriage' in (6.4c); (c) R arguments as in *pad ruwān* 'to the soul' or *pad šoy* 'to the husband' in (6.3a) and (6.3b) above.

The examples so far show that while the three distinct senses of Middle Persian *dād,* i.e., creation, caused change of location, and transfer, are sometimes expressed by distinct argument frames, they are also found expressed by identical argument frames. The reinterpretation of the different meanings of Middle Persian *dād* as senses of a single poly-functional verb is not surprising. It has been already amply argued in the literature that events of transfer are a specific type of caused change of location (see, for example, Jackendoff 1983, DeLancey 2000). Thus, speakers of Iranian in general have encountered two identical verb forms expressing notionally related event types, and it seems to be only a matter of time, then, until the same argument frames will express these distinct types of events as well.

A very similar situation is found in languages which use cognates of Proto-Iranian **daH* to express "beat, hit, strike"; indeed, Cheung (2007) cites "hit, strike" as one of the basic meanings of **daH*, and several sources cite such a meaning, or usage, for cognates of **daH*. For example, in his etymological vocabulary of the Shughni group (Eastern Iranian), Morgenstierne (1974:30) cites "to give, to strike etc." as the meanings attached to Shughni cognates of Proto-Iranian **daH*, but does not give any examples. It is argued here, however, that at least sometimes, the use of cognates of **daH* to express "beat, hit, strike" is synchronically related to the use of cognates of **daH* to express transfer.

In the data analyzed here, Gorani cognates of **daH* express both transfer and "beat, hit, strike" semantics. This is illustrated in (6.5-6.6) below. Examples (6.5a-b) express transfer in a dittransitive clause. In examples (6.6a-b) the Gorani cognate of **daH* is used in clauses expressing "beat, hit".

(6.5a) *šīr-aka=š ma-tī=ya wan=šān*
milk-DEF=3SG IND-give.PRS.3SG=DRCT to=3PL
'(and) she gave them the milk' (Gorani, Mahmoudveysi et al. 2012, 2:28)

(6.5b) *min nān ma-ti-m=a Alī*
1SG bread IND-give.PRS-1SG=DRCT Ali
'I gave bread to Ali' (Gorani, Mahmoudveysi et al. 2012:52, glosses mine)

(6.6a) *awał kām=mān bi-tī-yām wa kām=mān*
first which=1PL SBJV-give.PRS-1PL to which=1PL
'Which of us should strike which of us first?' (Gorani, Mahmoudveysi et al. 2012, 2:84)

(6.6b) *m-āy* *ma-tī=ya* *bara-ka*
IND-come.PRS.3SG IND-give.PRS.3SG=DRCT door-DEF
'He comes, he knocks on the door' (Gorani, Mahmoudveysi et al. 2012, 2:43)

The main difference between (6.5a-b) and (6.6a-b) seems to be verbal valence. In examples (6.5a-b), "give" is accompanied by three participants: the giver, the theme, and the recipient. In (6.6a-b), it is accompanied by two participants: the beater and the beaten element. The beaten argument in (6.6a-b), however, is expressed not as a prototypical P argument, but as a recipient, or R, argument. In (6.6a-b) the verb is followed by the directional marker, and the beaten participant is expressed by a postverbal phrase, and in (6.6a) it is also flagged by *wa(n)* 'to, for'.

The main syntactic difference between clauses expressing transfer or "beat, hit, strike" with cognates of ***daH,* then, seems to be the absence of an overt theme (T argument) in clauses expressing "beat, hit, strike" as in (6.6a-b). There are, however, several instances in the Gorani data where clauses expressing "beat, hit, strike" type events are accompanied by an overt T argument expressing the type of interaction of the hitter, or beater, and the beaten element. In (6.7a), the T argument is the horns of the goat; in (6.7b) it is an onomatopoeia for "knock". For the most part, when this happens, the beaten argument is often flagged by *bar* 'front, in, at'. This preposition is often used to flag goals of motion. Thus, there is at least some different in the coding of prototypical R arguments and the beaten participant in clauses with three arguments.

(6.7a) *tā min ʃāx=im bi-tī-m=a bar gaya=y gurg*
until 1SG horn=1SG SUBJ-give.prs-1SG=DRCT to belly=LNK wolf
'so that I thrust my horns into the belly of the wolf!' (Gorani, Mahmoudveysi et al. 2012, 2:76)

(6.7b) *taq taq ma-tī=ya bar ka=y pādšā*
knock knock IND-give.PRS.3SG=DRCT to house=LNK king
'He knocks on the king's house' (Gorani, Mahmoudveysi et al. 2012, 3:66)

In at least some of the instances where "give" expresses "beat, hit, strike", there seems to be a contextually recoverable zero in the T argument function. This is illustrated in (6.8), where the wolf picks up a stone, brings it over to the house of the goats, and uses it to bang on the door. The stone, which is clearly used to knock on the door, and would be the T argument in that clause (equivalent to the horns in (6.7a) above), is not overtly mentioned in the clause and can be contextually retrieved.

(6.8) *ya kuč-a qawī m-ār-ē ma-tī=ya wa bar*
INDEF stone-NA hard IND-bring.PRS-3SG IND-give.PRS.3SG=DRCT to door
'(He) brings back a hard stone, (and) bangs on the door' (Gorani, Mahmoudveysi et al. 2012, 2:49)

The examples in this section so far demonstrate that the grammar associated with cognates of Proto-Iranian **daH*, when it expresses "beat, hit, strike" events, is distinct from the syntax of transitive clauses and is more similar to that of ditransitive clauses. The beater or hitter participant is expressed as an A argument; the beaten participant is expressed as an R argument. The T argument expresses the type of contact the beater or

hitter participant has with the beaten participant ("knock, knock", "blow"), or the instrument used to make this contact ("stone"). In the Gorani texts published in Mahmoudveysi et al. 2012, such T arguments are often expressed by a zero mention. The zero mention can be contextually identifiable zero, as in (6.8), but it is also found when the type of interaction between the beater and the beaten participants is the prototypically expected one (e.g., knocking on a door).

Apart from "beat, hit"-type events and transfer events, the Gorani cognate of Proto-Iranian **daH* is also found in clauses expressing caused change of location, which correspond to the "place, put" function described for Middle Persian above. Examples (6.9a-b) illustrate this usage, and both the events described involve a violent interaction between two participants which ends with one being on the ground. The coding of the arguments in (6.9a-b) is identical to the one found in clauses expressing transfer events with the Gorani coganate of Proto-Iranian **daH*.

(6.9a) *bāwk=aš* *ma-tī=ya* *zamīn=ay*
father=3SG IND-give.PRS.3SG=DRCT ground=DEF
'(He) threw his father to the ground' (Gorani, Mahmoudveysi et al. 2012, 5:144)

(6.9b) *dā=š* *wa zamīn*
give.PST=3SG to ground
'(He) threw him to the ground' (Gorani, Mahmoudveysi et al. 2012, 4:136)

For the most part, then, the Gorani cognates of Proto-Iranian **daH* are found in ditransitive clauses, regardless of the type of event they express. The only unique coding property, differentiating one meaning of Gorani "give" from the others is the flagging of

the beaten participant by the preposition *bar* "on, at" in some clauses expressing "beat, hit, strike". The second main difference between the different uses of cognates of **daH* in Gorani is, as noted briefly above, that T arguments in clauses expressing "beat, hit, strike" are often expressed by a zero mention, either contextually identifiable or not, while the T arguments in clauses expressing transfer or change of location are often expressed by NPs, either lexical or pronominal. The total number of clauses expressing "beat, hit, strike" with the Gorani cognate of **daH* is rather small, and does not allow for any quantitatively informative analyses.

To summarize, the Proto-Iranian segment **daH* has been argued by Cheung (2007) to express six distinct homophonic verbs. Of these, three have survived mostly as lexical nouns, but three have survived as verbal lexemes expressing events of transfer, caused change of location, creation, and "beat, hit, strike." In many Iranian languages, where cognates of **daH* are found in clauses expressing these three different meanings, their grammar converged, and despite the fact that diachronically these different uses were an instance of homophony, synchronically cognates of **daH* are best described as polyfunctional or polysemous rather than as homophonous.

6.2 "Give" as Main Verbal Predicate

The rest of this chapter discusses the construction types in which cognate "give" verbs are found throughout the Indo-Iranian languages analyzed. This section concentrates on "give" verbs in construction types in which they function as the main verbal predicate. The different construction types described in this section are distinguished by the number of elements associated with "give", their coding properties, and their semantic role in the event described in the clause.

6.2.1 Monovalent and Copular Construction Types

Across the data, "give" is mostly found in clauses expressing trivalent events and, less frequently, bivalent events. In Palula, however, "give" is also found in copular clauses and in clauses expressing some of the six nominal predication functions.

In (6.10) below, "give" is the only verbal predicate in a clause expressing the proper inclusion function: i.e., the referents of *áa insaán* 'a person' belongs to the set denoted by *ɣaribí* '(the) poor'. Thus, "give" is used here in a clause expressing an inchoative nominal predication function. In his grammar of Palula, Liljegren (2016:204) argues that *de,* the suppletive past-tense form of "be, become," is a grammaticalization of a participle form of "give." The use of "give" in (6.10), Liljegren argues, may be a precursor to its grammaticalization as a suppletive past-tense copula.

(6.10) *áa insaán ɣaribí dít-u seentá*
INDF person poor give.PFV-MSG COND
'when a person becomes poor' (Palula, Liljegren & Haider 2015:87)

"Give" is also found in Palula in clauses expressing weather events such as snowing. In (6.11), "give" is the only verbal predicate in the first part of the sentence. The verb indexes *kir* 'snow'.[54]

(6.11) *dhoóɣ índa kir dít-u de típa bi kir hín-u*
yesterday here snow give-PRF.MSG PST now also snow be.PRS-MSG
'It was snowing here yesterday, and even now there is snow' (Palula, Liljegren 2016:176)

6.2.2 Bivalent Construction Types

6.2.2.1 With A and P (T) Arguments.

Cognate "give" verbs are often used in clauses expressing the creation of a new object. Here, the creator participant is coded as an A argument and the created participant is coded as a P (or T) argument. This construction type has been shown above for Middle Persian, where the creator participant tends to be divine. Example (6.12), repeated from (6.1a), illustrates this construction type. This is further illustrated in (6.13a-b) where cows create milk and butter without an intended recipient.

(6.12) *cē ka=m gēhān dād ā=m mēzd=ēw be kard*
for when=1SG world give.PST.3SG then=1SG offering=INDF PRV do.PST.3SG
'for when I created the world, I made an offering' (Middle Persian, PRDD 16.b)

[54] This is a possible Turkic loanword; compare Turkish, Kazakh, Kyrgyz, *kar* 'snow' and Uzbek *qor* 'snow' *inter alia* with Sanskrit *hima* 'snow' and Persian *barf* 'snow, frost'.

(6.13a) *deʃĩ gā-yẽ to na dud^{h} dẽ*

local cow-PL TOP NEG milk give.OPT.3PL

'Local cows don't give milk' (Hindi, Premchand 2017[1936]: 1.7)

(6.13b) *de put roɣan hā-din*

two put butter PRV-give.PRS.3PL

'They [a hundred cows] should give two put [unit of weight] of butter' (Ziyarat Mazandarani, Shokri et al. 2013:211)

It is possible that the Iranian occurrences of this construction type are the result of the loss of the phonological distinction between "put, place" and "give", following the loss of PIE aspirated stops. Example (6.13a), however, is from Hindi, an Indo-Aryan language, and similar examples found in Palula and Kotia Oriya show that the extension of "give" to be used in this construction type can also be the result of normal processes of semantic extension, since across Indo-Aryan, the distinction between PIE "put, place" and "give" was preserved.

6.2.2.2 With A and R Arguments

Infrequently in the data, "give" is accompanied by two arguments coded as an A and an R argument. This was illustrated in examples (6.6a-b) above, one of which is repeated in (6.14). These examples usually involve some kind of "hitting" semantics. In (6.14), the A argument is not overtly mentioned in the clause, but it is indexed on the verb. The participant being hit is expressed in (6.14) by a post-verbal NP, a relative position usually reserved in Gorani for R arguments and goals of motion. Other instances

of this construction type, such as (6.6a) above, involve flagging of the stroke participant by *wa* ‘to’.

(6.14) *m-āy* *ma-tī=ya* *bara-ka*
IND-come.PRS.3SG IND-give.PRS.3SG=DRCT door-DEF
‘he comes, he knocks on the door’ (Gorani, Mahmoudveysi et al. 2012, 2:43)

As noted above, this use of “give” across Iranian seem to be synchronically and diachronically related to ditransitive uses of “give” in which it is also accompanied by an overt T or P argument expressing the tool with which the agent will strike the referent of the R argument or the type of strike involved.

6.2.2.3 With a Deed Element and an A Argument

In this construction type, “give” is accompanied by an element which encodes the type of event the clause expresses, and an A argument expressing the sole participant in that event. In (6.15) from Palula, “give” is used together with two arguments. The first, the A argument “the women”, is not overtly expressed in the clause but is clear from the immediately preceding context, and it is also indexed on the verb. The P argument, *rhoóa* ‘songs’ is overtly expressed in the clause. While there was obviously someone present listening to the songs, they are not overtly mentioned in the clause, and as the listeners are not a specific person, the lack of an overt R argument here cannot be argued to be an instance of a contextually recoverable zero.

(6.15) *kuɽíina ta támbul-am-ii dʒe-en rhoó-a de-en*
woman.PL DS drum-PL.OBL-GEN beat-3PL song-PL give-3PL
'The women beat the drums and sang songs.' (Palula, Liljegren & Haider 2015:63)

6.2.3 Trivalent Construction Types

6.2.3.1 Transfer Events

Across the Indo-Iranian languages analyzed here, "give" is used in several different trivalent construction types expressing transfer events in which an agent (A argument) transfers a theme (T argument) to a recipient (R argument). The construction types identified here differ in the coding properties of their T and R arguments. That is, these construction types differ in their ditransitive alignment. The various alignments found here are not limited to crosslinguistic variation, but there is also intralinguistic variation, either in what might be thought of as ditransitive alignment splits or as a result of some alternation. Further, ditransitive alignment seems to interact with differential object marking (DOM) in a number of ways, which leads to situations which I analyze as instances of split alignment. Thus split ditransitive alignment with "give" seems to be the norm across Indo-Iranian. The goal of this section is not to provide meticulous documentation and analysis of this variation, but rather to illustrate this typological variation with examples from different languages across the family.

The most common transfer construction type with "give" is the indirective alignment type. In this, T arguments are coded as P arguments and R arguments are coded differently. Across the data, T and R arguments are often distinguished by their flagging, their interaction with verbal indexing, and their relative word order with the

verb. In examples (6.16a-b), the T argument is expressed by a morphologically unflagged NP and the R argument is flagged by an adpositional marker. In (6.16a), from Palula, the T argument is expressed by an unflagged NP and the R argument is flagged by the post-position *the* 'to'. In (6.16b), a Gorani example elicited by Mahmoudveysi et al. (2012:52), the T argument is morphologically unflagged and the R argument is flagged by the preposition *wa* 'to'.

(6.16a) *ma tas the páandʒ sáu rupeé baxʃíʃ de-án-u*
1sg.NOM 3SG.REM.ACC to five hundred rupees reward give-PRS-MSG
'I would give him a five-hundred-rupee reward.' (Palula, Liljegren & Haider 2015:144)

(6.16b) *min wa Alī nān ma-ti-m*
1SG to Ali bread IMPF-give-1SG
'I give Ali the bread.' (Gorani, Mahmoudveysi et al. 2012:52)

In (6.16a-b), the marker on the R argument is employed to flag (human) goals. There are, however, instances where this marker is locative in nature. This is shown in (6.17a), from Middle Persian (repeated from (6.3a) above), where the locative preposition *pad* 'at, in, by' flags the recipient. In (6.17b), from Ziyarat Mazandarani, the post-verbal R argument is flagged by *de*, another locative marker.

(6.17a) *duxt ... pidar pad šōy be dah-ēd*
daughter …father in husband PRV give.PRS-3SG
'(his own) daughter the father gives to this husband' (Middle Persian, RAF Q62)

(6.17b) *masan de se=tā he-dā-yim in vače=mun=de*
for.example two three=CLF PRV-give.PST-1PL DEM child=1PL=LOC
'For example, we gave two or three to our child.' (Ziyarat Mazandarani, Shokri et al. 2013:117)

In some languages in the sample, especially modern Central Iranian languages, the relative order of the T argument, the R argument, and the verb also distinguishes the coding of the T vs. R arguments. This is shown in (6.18), an elicited Gorani example (Mahmoudveysi et al. 2012). Here the R argument, Ali, is expressed by a post-verbal unflagged NP. In the data from Central Iranian languages analyzed in this study, a large portion of overt R arguments are expressed by such post-verbal NPs. This tendency is correlated with the tendency of goals of motion to be expressed by a post-verbal NP.

(6.18) *min nān ma-t-im=a Alī*
1SG bread IND-give.PRS-1SG=DRCT Ali
'I gave Ali the bread.' (Gorani, Mahmoudveysi et al. 2012:52)

This difference in relative word order is often accompanied by a difference in the flagging of the T and R arguments. In (6.19a), from Sivandi, the R argument is expressed by a post-verbal NP and the T argument is flagged by the differential object marker. In (6.19b), from Gorani, the R argument is post-verbal and is flagged by *wa* 'to', the same preposition that flags it in (6.19b) above.

(6.19a) *xūī=ra be-de-y ossād*
egg=OBJ IRR-give.PRS-1SG master
'I will give the egg to the master.' (Sivandi, Lecoq 1979:85)

(6.19b) *bizbal šīr ma-t-i=ya wa dāya kaywānu*
Bizbal milk IND-give.PRS-3SG=DRCT to mother woman
'Bizbal gives milk to the old lady.' (Gorani, Mahmoudveysi et al. 2012, 1:100)

Transfer events are also expressed in the data by a construction type which shows secondative alignment. Here R arguments are flagged by the same markers that flag P arguments, while T arguments are expressed by morphologically unflagged NPs. This is demonstrated in (6.20a-b) below. In (6.20a), the R argument is flagged by *ke* and the T argument is expressed by the phrase *e laḍu* 'these sweets'. In (6.20b), the R argument is flagged by *re* and the T argument is expressed by *ʃelvār* 'trousers'.

(6.20a) *mo=ke e laḍu de-lai*
1SG=OBJ DEM sweet give-PST.3PL
'They gave me these sweets.' (Kotia Oriya, Gustafsson 1973a:186)

(6.20b) *yeki=re ʃelvār dā*
someone=OBJ trousers give.PST.3SG
'She gave someone trousers.' (Ziyarat Mazandarani, Shokri et al. 2013:77)

The different coding means of T and R arguments do not always correlate. In (6.20c), from Ziyarat Mazandarani, the object marker *=re* is found flagging a post-verbal R argument. Thus, the alignment of flagging is secondative, as the R argument is flagged as a P argument. In terms of relative word order, the R argument appears post-verbally, and thus is coded differently from the P arguments which are almost always pre-verbal.

(6.20c) *unvax in bard-e unje dā ārus=re*
now DEM take.PST-3SG here give.PST.3SG bride=OBJ
'He took it [the chador] and gave it to the bride.' (Ziyarat Mazandarani, Shokri et al. 2013:151)

Across the data, there are many constructions in which R arguments are flagged by the same markers that flag P arguments. This flagging occurs, however under slightly different conditions for each of the two argument types. Markers of P arguments in both Kotia Oriya and Ziyarat Mazandarani, for example, are deployed as differential object markers under specific discourse conditions which involve identifiability, specificity, etc. (see, for example, Shokri et al. 2013:23).[55] When the same markers flag R arguments, their deployment seems to be less motivated by discourse considerations and more by syntactic considerations, since R arguments are flagged by these markers regardless of their discourse status. Consider, for example, (6.21) below, from Kotia Oriya. In this example, the R argument expresses a general, non-specific and non-identifiable referent "some other person". It is, however, flagged by the Kotia Oriya Object marker *=ke*.

(6.21) *ar goţek lok=ke de-la*
again INDEF person=OBJ give-PST.3SG
'He gave [some of these sweets] to another person.' (Kotia Oriya, Gustafsson 1973a:186)

[55] Shokri et al. 2013:23, and many others, consider the *=re* / *=ra* markers on P arguments to be a borrowing from New Persian. Stilo (2004) seems to point to a sociologically stratified use of this marker in Gazi, correlated with education and exposure to New Persian. The question, then, is what motivated the deployment of this marker to flag R arguments, which is not one of its functions in Standard New Persian.

Even when R arguments are overtly flagged by the same marker as P arguments, the conditions governing the flagging are not always identical. The T argument in such ditransitive clauses are often expressed by an unflagged NP or a zero mention (see (6.20a) and (6.21) above, from Kotia Oriya). In some languages, however, the T argument is also flagged by the object marker under the same DOM conditions as a P argument. In such clauses, then, the T argument is sometimes coded identically to the R argument, and both are coded identically to the P argument. Thus, these clauses have a double object alignment.[56] This is illustrated in (6.22a-b) below, from Kupia. First, in (6.22a), the T argument is expressed by the unflagged NP *e:ku bo:da* 'one child' and only the R argument is expressed by the 1PL pronoun flagged by *ka* (cognate of the Kotia Oriya *ke*). In (6.22b), we have double object marking: the R argument is expressed by the same NP as in (6.22a), but now the T argument is expressed by the (nominalized) adjective *ḍakilo* '(the) small (one)', referring to the smaller of two children. As this T argument is specific and identifiable, it is flagged by *ka*.

(6.22a) *e:ku bo:da am=ka de*
one child 1PL=OBJ give.IMPR.SG
'Give one child to us.' (Kupia, Christmas & Christmas 1973a:54)

(6.22b) *ḍakil-o=ka am=ka de*
small-MSG=OBJ 1PL=OBJ give.IMPR.SG
'Give us the small one,' (Kupia, Christmas & Christmas 1973a:62)

[56] It might be beneficial to consider such constructions as members of a secondary double-object split ditransitive alignment motivated by identifiability and specificity of the T argument.

An identical flagging of T and R arguments is also found in Middle Persian, where both can be expressed by morphologically unflagged NPs. This is illustrated in (6.23a-b). In (6.23a), the T argument is expressed by "thing" and the R argument by "the pious". In (6.23b), the T argument is expressed by "milk" and the R argument is expressed by "her own child", both unflagged NPs.

(6.23a) *kār ud kirbag was kun-ēd ud ahlaw-ān čiš dah-ēd*
work and merit much do.PRS-3SG and pious-PL thing give.PRS-3SG
'He does many good deeds, and give things to the pious.' (Middle Persian, PRDD 28)

(6.23b) *ēn ruwān ī awēšān zan-ān*
DEM soul LNK 3PL woman-PL
ke kōdak ī xwēš šīr ne dād
REL child LNK REFL milk NEG give.PST.3SG
'These are the souls of these women who didn't give milk to their children.' (Middle Persian, AWN 95.3)

One could argue that because in Middle Persian the verb in the past tense indexes the P argument, the indexation in constructions like the one in (6.23b) could potentially discriminate between the T and the R arguments. This is true, but across the data analyzed here all instance of different T and R arguments found in this Middle Persian construction would motivate the same form of the verb. In (6.23b), for example, both the T and the R arguments are the 3rd person singular, and both could be argued to be indexed on *dād* 'gave'. Furthermore, in clauses like (6.23a) the verb is in the present tense and indexes the A argument. Such clauses cannot be used to differentiate the coding

properties of T and R arguments. Hence, (6.23a-b) can be argued to be an instance of a double-object alignment type.

Another transfer construction type in which "give" is attested in the data expresses payment or a commercial, often barter, transaction. In this construction type "give" is accompanied by the giver, or payer, participant, coded as the A argument, and the amount paid is coded as the T argument. The recipient of the money, i.e., the seller, is not overtly expressed in this construction type, and is often understood from context. This construction type is illustrated in (6.24a-b) below. Example (6.24a), from Hindi, discusses a potential transaction in which a nice-looking milking cow changes owners. The speaker, trying to express how much the cow could be sold for, says that some people would pay a large amount for it. Examples (6.24b), from Bactrian, is taken from a contract.

(6.24a) *acch-i raqam dete the*
good-F amount give.PTCP be.pst.3SG
'(they) would give a nice amount (for it)' (Hindi, Premchand 2017[1936]: 1.8)

(6.24b) *taδo=ieːio oauaɣo na laδeːio*
then=3SG price NEG give.PST.OPT.3SG
'then he did not pay the price' (Bactrian, Sims-Williams 2008, ed8)

The goods or services which are being sold in the events described in clauses like (6.24a-b) are often contextually identifiable but are rarely overtly mentioned in the clause. When they are mentioned, they are usually coded as the "indirect affectee" macrorole. The flagging of such a participant is sometimes identical to the flagging of

P/T arguments, especially in Iranian. This is shown in (6.25) from Ziyarat Mazandarani, describing a barter in which oxen were lent for ploughing rice fields. The A argument in (6.25) is the land owners, giving the forty *man* of rice. The goods they got in exchange, the oxen, are overtly mentioned and are flagged by *=(r)e,* the Ziyarat Mazandarani Object marker. One function of the final *=(r)e* is to express the macro-role of indirect affectee (Shokri et al. 2013:24), and this seems to be its function here.

(6.25) *in varzā varzā=e čel man šāli dā haryak varzā.*
DEM ox ox=OBJ forty man paddy.rice give.PST each ox
'For each ox, they gave forty *man* of paddy rice.' (Ziyarat Mazandarani, Shokri et al. 2013:211)

The final transfer construction type identified with "give" in the data is when "give" is accompanied by a non-finite form of a verb which indicates what the referent of the R argument has to do with the given theme. The T argument, however, is rarely overtly expressed in the clause. Importantly, the infinitive in such clauses, without an overt marking of its own arguments, is dissimilar from the usual expression of purpose clauses across the construction family.

(6.26a) *vāʃ yārd-im dāy-im gu=ā xwārd-an*
vash.plant bring.PST-1PL give.PST-1PL co=OBJ eat-INF
'we brought vash plants and give (them) to the cows to eat' (Ziyarat Mazandarani, Shokri et al. 2013:65)

(6.26b) *u=š ō xwardan mardomān dād*
and=3SG to eat-INF person-PL give.PST.3SG
'and gave (it) to people to eat' (Middle Persian, AWN 84.3)

(6.26c) *aru tae-laga manu-khan=ke bi kha-bole di-se.*
and 3SG-GEN man-PL=OBJ ADD eat-PTCP give-PST
'and he gave some to his companions to eat' (Nagamese, Luke 6:4)

6.2.3.2 Giving an Event

In several ditransitive construction types in the data, "give" is used together with a nominal element encoding the type of event or situation expressed by the entire clause as a whole. Following the terminology used in this dissertation, these nominal elements are referred to as the deed elements. In the construction type reviewed in this subsection, "give" and the deed element are accompanied by two additional NPs that express the participants in the event encoded by the deed element. These construction types seem like good candidates for N-V complex predication status, but the discussion here remains agnostic in this question.

In this construction type, "give" is accompanied by a deed element and two additional NPs coded as A and P arguments. This construction type is illustrated in (6.27a-b) from Nagamese and Sivandi. In (6.27a) the deed element is *jonom* 'birth', the A argument is *apuni* 2SG, and the P argument *etka chokora* 'one child' is flagged by the Nagamese Object marker. This example seems like a calque from English, the official language of the Indian state of Nagaland where Nagamese is spoken: i.e., *give birth*. In (6.27b), from Sivandi, the deed element is *tow* 'turn', the A argument is expressed by the

first person singular pronoun, and the P argument is expressed by *gūsala* 'ram', flagged by the differential object marker.

(6.27a) *apuni ekta chokora=ke jonom di-bo*
2SG one child=OBJ birth give-FUT
'You will give birth to a child.' (Nagamese, Luke 1:31)

(6.27b) *me gūsala=rā tow me-de-y*
1SG ram=OBJ twist IMPF-give.PRS-1SG
'I will turn this ram.' (Sivandi, Lecoq 1979:109)

There is, however, a difference between the two constructions in (6.27a-b). The coding properties of the deed and the A and P arguments in (6.27a) are identical to those found in the usual Nagamese ditransitive transfer constructions, where the R argument is flagged by *ke* and the T argument is unflagged. Thus, there is no evidence in clauses like (6.27a) that the deed element motivates overt changes to valence and argument structure.

In example (6.27b) from Sivandi, in contrast, *gūsala* 'ram' is flagged by *rā*, the differential object marker in Sivandi. Across the Sivandi data analyzed, R arguments are not attested flagged by *rā*, but are expressed by mostly postverbal NPs flagged by the preposition *ba*. Thus, *gūsala* 'ram' is the P, not R, argument in (6.27b), and there is no other candidate to fill the function of an R argument. In such examples, then, the deed

element can be argued to motivate a change in the argument structure of the entire clause.[57]

Many instances of "give" across the data are ambiguous between a transfer event reading and a reading in which the T argument is an event. This is illustrated by (6.28a-b) from Ziyarat Mazandarani. In (6.28a), the NP *šir* 'milk' could be taken to literally mean 'milk' so that the clause expresses a transfer event, or it could be taken to encode the event of breastfeeding. Similarly, *telāɣ* in (6.28b) could be taken either to mean 'divorce papers' or to encode the event type of divorcing.

(6.28a) *har de vačā=re šir dā*
all two children=OBJ milk give.PST.3SG
'She breastfed both children.' (Ziyarat Mazandarani, Shokri et al. 2013:218)

(6.28b) *nā hiʃki zan telāɣ he-dā*
NEG no.one woman divorce PRV-give.PST
'No, no one divorced their wife.' (Ziyarat Mazandarani, Shokri et al. 2013:78)

When the deed element in the clause encodes a PCU event type, the clause often (but not always) includes a complement clause expressing the contents of the PCU event. This is shown in (6.29a-b), where "give" is followed by a complement clause. In (6.30a) the complement clause includes a subjunctive main verb and is thus an indirect quote of

[57] Thus, by the criterion used in Chapter VIII for N-V complex predication, the joint determination of argument structure and valence by the deed element and the verb (following Mohanan 1997 and many others), there is enough evidence to consider (6.27a) an instance of complex predication.

the permission given. In (6.29b) the complement clause is basically a direct quote of the contents of the speech.

(6.29a) *baʕd eǰāza ma-tī=ya ka bi-nīš-ām*
later permission IND-give.PRS.3SG=DRCT COMP SUBJ-sit.PRS-1PL
'Then they gave (us) permission to sit (with them).' (Gorani, Mahmoudveysi et al 2012, 7:11)

(6.29b) *tíi eeɽé baát dít-i*
3SG.REM.OBL DIST speech give.PRF-FSG
ki anú íɳţ-a sangí mháala …
COMP DEM bear-OBL with wrestling …
'He said (he would give 500 rupee) to whomever would wrestle the bear.' (Palula, Liljegren & Haider 2015:144)

6.2.4 "Give" in Change of Location Clauses

The final construction type identified in this subsection is one in which "give" is used alongside three nominal elements in clauses expressing caused change of location. In such constructions, the new location is usually encoded by a locative PP, the agent is coded as an A argument, and the undergoer is coded as a P argument. This construction type has already been illustrated for Middle Persian and Gorani. In example (6.30) from Middle Persian, repeated here from (6.1b), the locative phrase is headed by *pad* 'in, at, by':

(6.30) *ud man āxist h-om u=m pad zamīg be dād*
and 1SG rise.PST be.PRS-1SG and=1SG in earth PRV do.PST.3SG
'and I rose, and put him (Wind) on the earth' (Middle Persian, PRDD 18)

It could be suggested that this construction type is attested in Iranian as a result of the neutralization between PIE 'put, place' and 'give', which resulted in the homophonous root **daH* in Proto-Iranian. Across Indo-Aryan, however, one also finds similar usages of "give" as the main verbal predicate in clauses expressing caused change of location. Two such examples are given below, from Palula and Hindi. In (6.31a) the locative phrase is flagged by the Palula post-position *wée* 'into'. In (6.31b), the locative phrase is flagged by *mẽ,* the Hindi general locative marker.

(6.31a) *se lhaást-í=wée de-í asaám be dharíit-a*
DEF plain-OBL=into give-CV 1PL.ACC 1PL.NOM remain.PFV-MPL
'We were carried away to that flat place and came to rest there.' (Palula, Liljegren & Haider 2015:134)

(6.31b) *bʰola=ne gay=ki pagahi-yā hori=ke*
bhola=ERG cow=GEN.FSG tether- PL hori=GEN.MSG
hātʰ=mẽ dete hue
hand=LOC give.PTCP be.PRS.3SG
'Bhola put the tether of the cow in Hori's hand.' (Hindi, Premchand 2017[1936]: 1.17)

6.3 "Give" in Serial or Serial-Like Constructions

Across the languages analyzed here, "give" is found in monoclausal constructions in combinations with other verbs to denote what seems to be a single, complex event involving transfer and some distinct related aspect of an event (e.g. "buy", "bring"). The two verbs in this construction type share TAM and polarity, and tend to share most, if not all, participants. The discussion of the different properties of this construction type could

easily be the subject of a monograph or a series of dedicated papers. Here I aim only to establish this as a distinct construction type, attested in several languages of the family, and show that it is quite similar to what has been described as serial verb constructions in other languages.

This construction type is shown in (6.32a-b). The verb "give" is preceded by "buy" and "bring", and the two verbs share the agent and the theme, the horse in (6.32a) and the thread in (6.32b). The R arguments in both examples are licensed by the valence and argument structure of the "give" verb but not by the "buy" or "bring" verbs. Nevertheless, they are expressed before both verbs in each example and not between the two or after the "give" verb. This pattern is especially notable in Ziyarat Mazandarani, a Central Iranian language, in which post-verbal goals and recipients are common. Thus, there is at least some positive evidence for some type of higher cohesion between the two verb forms in (6.32a-b).

(6.32a) *am=co bo:da=ka i:njo go:ḍo gen-a da-a gel-t-a*
1SG=GEN child=OBJ DEM horse buy-NF give-NF complete-PRS-1SG.OPT
'I would have bought this horse for my son.' (Kupia, Christmas & Christmas 1973a:73)

(6.32b) *kārvefā=mān nax ba-vr-em hā-de-m*
worker=1PL thread SUBJ-bring.PRS-1PL PRV-give.PRS-1PL
'(If no one had them) we would bring threads and give (them to) our workers.' (Ziyarat Mazandarani, Shokri et al. 2013:222)

The high degree of cohesion between "give" and "bring" or "buy" in (6.32a-b), and in many other clauses across the data, suggests that these clauses represent a usage of

"give" which is qualitatively different from its use as a finite verb on the one hand, since it is tightly integrated with another verb, and from its use as an auxiliary or compound verb on the other hand since it expresses events of transfer here, and not grammatical categories such as TAM.

It is tempting to describe the construction type in (6.32a-b) as a serial verb construction. There are, however, several issues with using the term "serial verb" here. First, many definitions of serial verb constructions are aimed at capturing some phenomenon in a specific language or a language group, either genealogical or areal, and hence make use of language-specific constructions and facts. As the languages analyzed here all form a coherent genealogical group, one could potentially use the same approach for defining serial verbs. The problem with such an approach is that the potential serial verb constructions across Indo-Aryan are quite different syntactically. For example, in (6.32a) the serialized verbs are both non-finite, while in (6.32b) both are finite (but share the same TAM and index the same participant).

In order to identify serial-verb or serial-like constructions in this study, I have opted to follow the characterization of serial verbs given in Haspelmath 2016 which attempts to define serial verbs as a "comparative concept", avoiding language-specific constructions and properties. Haspelmath 2016 defines serial verb constructions as "monoclausal constructions with two independent verbs without an overt element linking them and with no predicate argument relationship between the verbs" (2016:292).

Much of these criteria correspond nicely to what we find in examples like (6.32a-b) above and many other examples in the data. First, the combination of the two verbs is not a lexicalized or idiomatic expression; instead, each verb donates its own lexical

semantics to the entire clause. In (6.32a) there are two subevents, the first being the purchase of the horse and the second the transfer of the horse to the son of the speaker. In (6.32b) there are also two subevents, bringing the thread, and giving it to the weavers, who, in that context, had run out of thread.[58]

Examples of this construction type are monoclausal. As Haspelmath notes, the criteria for whether a specific expression is a single clause or two clauses are highly language-specific. To briefly give some language-specific evidence that suggest that examples such as (6.32a-b) above are monoclausal, consider the following points. First, the two subevents encoded by the two verbs share at least two participants, the agent and the theme. The recipient participant, however, pertains to the valence of the "give" verb and is not motivated by the first verb in the construction. It coded like an R argument and is expressed in a position preceding both verbs. This implies that the R argument is somehow treated as part of the combined argument frame of both verbs together in the construction and not to "give" alone. The R argument is never found in a position which delineates it from the non-"give" verb (i.e., between the two verbs).

Furthermore, both verbs in this construction type always have the same TAM and polarity categories. In (6.32a) above, both verbs are in their basic non-final or converbal form, and the auxiliary verb *gel* 'finish, complete', which encodes completive semantics,

[58] The degree to which the two verbal slots in the construction can be filled by different verb forms is a difficult to determine. When "give" is the second verb in the construction, the first verb usually has some semantic aspect of obtaining an object (by the giver): "buy", "bring", "take", "pick up", "find" etc. Thus, while several distinct verbs are attested in this construction, their semantic range does seem to be somewhat limited.

has scope over both. In (6.32b), both verbs are finite, both index the same participant (the A argument of both), and both have subjunctive or irrealis semantics.[59]

In Kupia, discounting cases involving an auxiliary, only the second verb will be overtly marked for TAM and person, and these have scope over the entire clause including the first verb in the construction. This is illustrated in (6.33), where "bring" is in the participle form and "give" is in the Kupia Imperative Singular form. Both verbs, however, have imperative semantics:

(6.33) *aŋ=ka donni amb-a a:n-a de*
1SG=OBJ two mango-PL bring-PTCP give.IMPR
'"Bring me two mangos," she said.' (Kupia, Christmas & Christmas 1973a:37)

The first two criteria used by Haspelmath to characterize serial verb constructions seem to correspond to the construction type described in this section. It is less clear, however, how well the criteria of an independent verb and lack of overt linking elements fit this construction type. Haspelmath's characterization of the verbs as "independent"[60] seems to mean that the verbs need to be able to express dynamic events without any additional inflectional coding. This criterion seems to be aimed at excluding aspectual auxiliaries and other types of grammaticalized functions of verbs. In examples (6.32 – 6.33) above, both verbs have a lexical and not grammatical function.

[59] One could argue that the subjunctive in (6.32b) is overtly expressed only once, on "bring". The issue with this claim is that in the Central and Western Iranian verbal morphology, the subjunctive preverbal clitic tends not to be overtly marked when another preverbal element (e.g., negation) is attached to the verb.
[60] Haspelmath goes as far as to present a comparative concept of "independent verb." (2016:303)

The verb forms in (6.32b), from Ziyarat Mazandarani, are fully finite and can clearly be used to express dynamic events without any further verbal morphology beyond what they already carry. Similarly, it is clear that in (6.32b) there are no additional grammatical linkage markers such as conjunctions between the two verbs. But a problem with describing instances of this construction type as a serial verb construction type arises across Indo-Aryan. In many Indo-Aryan languages, including Kupia and Kotia Oriya, the first, or non-final, verb in this construction (and many other constructions involving more than a single verb form) carries some non-finite morphology (i.e., it is not a "bare verb"), and is also not a fully finite conjugated verb. Such verb forms have been often described in the literature as "converbs" or "non-final" verb forms.

The two questions asked here in regard to the use of these forms in putative serial verb constructions is whether they are "independent" verb forms, as Haspelmath requires, and whether they can be analyzed as involving some indication of linkage. I will illustrate the difficulty with giving a clear and strict answer to these questions by using Kupia data, but a similar analysis can be offered for Kotia Oriya and other Indo-Aryan languages in which the first verb in the serial verb construction is not a morphologically bare verb.

The question of whether the Kupia non-final forms in *–a* can function as independent verb forms can be rephrased, following Haspelmath's criterion, as whether these forms can be used as the only verbal predicate in clauses expressing a dynamic event, without any additional grammatical markers (e.g., auxiliary verbs) and without ellipsis of such markers. The Kupia verb form in *–a* is often used in clauses an auxiliary verb (see the discussion of the auxiliary use of "do/make" in chapter IV; for its use in a compound verb construction, see subsection 6.4.2 below). It is also used, however, as the

only verbal predicate in the clause in chains of thematically and temporally closely-related events. This use of the Kupia *–a* verb form is illustrated in (6.34), where it is found twice with *wik* 'sell'. These verb forms are not overtly marked for TAM and person categories; instead, their TAM semantics are interpreted based on the fully finite verb *katasum* 'we eat' that occurs at the end of the chain.

(6.34) *roːju=ka akkuqu wik-a daːru wik-a aːmu ka-t-asum*
day=OBJ leaf sell-NF wood sell-NF 1PL eat-PRS-1PL
'We live (lit. eat), (by) selling leaves, selling wood day by day.' (Kupia, Christmas & Christmas 1973a:53)

The Kupia verb form in *–a*, which is always the form of the first verb in the construction type analyzed here, can be interpreted as an "independent" verb form because it can be the only verbal predicate in a clause without any additional marker beyond the *-a* suffix or ellipsis. The same verb form, however, can alternatively be interpreted as dependent, as it always depends on some other verb form, either in the same clause or in a different, subsequent clause, for interpretation of TAM categories and participants.[61]

Because the Kupia verb form in *-a* is always dependent on some other verb form for interpretation, one of the functions of the Kupia verb form in *-a* is to encode linkage between one verb form to a subsequent verb form or clause. It seems, then, that one cannot argue that the Kupia examples in (6.32a) and (6.34) above are clear instances of a

[61] One could further problematize this and ask whether the TAM categories are inherited (i.e., always identical) or interpreted (i.e., under some conditions the clauses may differ in TAM categories).

serial verb construction by Haspelmath's definition. This is similar to the situation in Japanese mentioned by Haspelmath (2016:304-305) based on Shibatani (2009), where a verb form with comparable range of usages to the Kupia verb form in *–a* is used in the putative serial verb construction.[62]

The goal of this section has been to establish the existence of cognate "give" verbs in a serial, or serial-like construction type, and to briefly describe the main properties of this construction. It has been shown here, albeit briefly, that such a use can be identified in both Iranian and in Indo-Aryan. Further, it has been shown that these uses conform largely to Haspelmath's 2016 characterization of serial verb constructions. Whether one analyzes the Kupia construction in *–a* as an instance of serial verbs (thus discounting the linking semantics of *–a*) or as an instance of "verb seriation" (as Haspelmath's 2016:305 suggestion), it is obvious that this is a functionally and structurally related usage of "give" across Indo-Iranian.

6.4 "Give" as Non-Main Verbal Predicate

In some of the Indo-Aryan languages analyzed here, "give" is not used as the main verbal predicate of the clause. Across the data, two such construction types have been identified. The first, called here the "let" auxiliary construction, is found across Indo-Iranian, and is also quite common elsewhere in West Asia (e.g., Modern Hebrew). The second, a V-V compound verb construction, is found only in new Indo-Aryan

[62] A similar situation is found across Turkic; see discussion in Johanson 1995.

languages. These constructions have received quite a lot of attention in the Indo-Aryan literature, and will be only briefly mentioned here.

6.4.1 "Let" Auxiliary

"Give" is often used in an auxiliary construction which involves permissive modality. This usage has been identified and described for Urdu by Butt (1995, 1997), and a very similar or identical construction is attested in Hindi. This construction type in Hindi and Gujarati is shown in (6.35a-b) below.

(6.35a) *ənu=ne sunīta=ko vəhã bhəjən nəhĩ̃ ga-ne diya*
Anu=ERG Sunita=OBJ there bhajan NEG sing-INF give.PRF.MSG
'Anu did not allow Sunita to sing devotional songs there.' (Hindi, Kachru 2005:205)

(6.35b) *kan-o loʈ-o pən mê ta-ra hath-mã*
hole-MSG jug-MSG even 1SG.ERG 2SG-GEN.SG hand-LOC
aw-wa nə di-dh-o
come-INF NEG give-PST-MSG
'I did not let even a jug with a hole come into your hand!' (Gujarati, Christian 1987:242)

In terms of participants, the auxiliary "give" verb adds to the overall valence of the clause the role of the person who is permitting. Thus, the overall structure of the clause is jointly determined by two verbs, the main lexical verb, "sing" in (6.35a) and "come" in (6.35b), and the "give" auxiliary. Based on this argument structure related property of this construction, Butt (1995, 1997) analyses this construction as an instance

of V-V complex predication (see Chapter VII for discussion of complex predication). In (6.35a-b) the participant added by the "give" auxiliary verb is coded as the A argument. The participants who are being given the permission are expressed as P arguments in (6.35a-b).

This construction type is very common across Indo-Aryan, but it has also been identified in some Iranian languages, as illustrated in (6.36). In this example, the A argument is the earth which is (not) granting the permission, and the person who would have undertaken the motion with permission is expressed as an oblique object with *wan* 'to him'.

(6.36) *zamīn řā ni=ma-tī-ya* *wan*
earth go NEG=IND-give.PRS-3SG to.3SG
'(he was so strong) the earth did not let him go' (Gorani, Mahmoudveysi et al. 2012, 5:101)

6.4.2 V-V Compound Verb

The final construction type is the well-known Indo-Aryan compound verb construction, which has received much attention in the literature (e.g., Hook 1975, 1991, 1995, Butt & Lahiri 2013, Poornima 2012). It is an innovation of the modern Indo-Aryan languages. (Massica 1991; Butt & Lahiri 2013 have argued against this position, but their argumentation has been well-countered in Slade 2013.)[63] The goal of this section is not to provide a full account of this construction type across the Indo-Aryan languages

[63] Hook (1991) cites two examples from Pali, one with "give" and the other with "put". In the texts analyzed for this study, which are composed of Jataka tales, there are no clear instances of this construction type.

represented in the data,[64] but rather to describe it briefly, and to describe the function of "give" in it.

The V-V compound verb construction is an auxiliary construction in which the first verb, often referred to as the main lexical verb, encodes the situation expressed by the entire clause, and the second verb, often called the "vector" verb, is usually one of a closed set of semantically bleached verbs such as "give", "take", "go", "come". In all Indo-Aryan languages in which V-V compound verbs are attested, "give" is one of the more prominent verbs in this function. This construction is shown in (6.37a-b) from Kotia Oriya and Gujarati, where the main lexical verbs are "build" and "teach", respectively. "Give" carries all the grammatical categories of tense and indexes the A argument.

(6.37a) *goʈek gor bin-e band-i de-lai*
INDEF house outside-LOC build-NF give-PST.3PL
'They build a house outside (for her).' (Kotia Oriya, Gustafsson 1973a:190)

(6.37b) *ta-ri widya tû koi=ne ʃikʰəw-i d-e*
2SG-GEN.FSG knowledge 2SG someone=OBJ teach-CV give-prs.2SG
'if you teach someone your knowledge' (Gujarati, Christian 1987:277-278)

The function of these constructions is usually described as two-tiered: the semantic contribution of the entire construction and the semantic contribution of the

[64] There is some variation across Indo-Aryan in both the frequency and the exact function of this construction. This has been briefly explored by Hook (1991, 1995) and Butt & Lahiri (2013, comparing Urdu and Bengali), but most analyses of this construction type concentrate on specific elements of its grammar in a particular language (e.g., Poornima 2012 for Hindi).

specific auxiliary ("vector") verb used. In Hindi, for example, the construction itself has been argued to express perfective aspect (e.g., Hook 1991) or event boundedness (Poornima 2012:199, who analyzes the function of this construction as signaling that an event has reached some normative endpoint[65]). These analyses, while differing in detail, are very close to one another, and the analysis proposed by Poornima can be seen as a more refined elaboration of the one presented by Hook.

The function of the entire V-V compound verb construction in other languages, is less clear than in Hindi. Hook (1991, 1995) notes that the Hindi Compound Verb construction is sensitive to abstract discourse features, while its Marathi and Gujarati counterparts mostly show sensitivity to their immediate clausal environment.[66] These differences between the distribution of this construction type are argued by Hook (1991, 1995) to represent different degrees of grammaticalization: he claims that in Hindi, where the construction is more sensitive to abstract semantic and discourse features, it is further along on the grammaticalization pathway than its Gujarati and Marathi counterparts.

The second tier in the semantics of the V-V compound verb construction has to do with the semantics associated with the choice of specific auxiliary or "vector" verb. As mentioned above in this subsection, across the Indo-Aryan languages in which this construction type is attested there are several different verbs attested in this function. These verbs tend to be more semantically abstract, or bleached, to begin with, and

---

[65] Poornima 2012:196-216 contrasts V-V auxiliary constructions with simple past tense verb forms which do not express that the event had reached a result state (see Pederson 2008 for Tamil, and Arunachalam and Kothari 2011 for Hindi)

[66] On the other hand, Christian 1987 argues that the Gujarati Compound Verb construction is associated with discourse functions such as fast-pacing the narrative main-event line and suddenness or unexpectedness of an event. While Christian's corpus is used in this study, his claims are not evaluated here (see Hook 1995 for an unfavorable evaluation).

include “give”, “take”, “put”, “go”, “come”. Hook (1975) suggests that in Hindi, the contribution of “give” in a compound verb construction has to do with signaling the fact that the action is performed for the benefit of someone other than the agent. Poornima 2012 demonstrates this function in Hindi by contrasting the use of “give” and “take” in the two constructed examples in (6.38a-b) below. Both clauses express the creation of a house by the agent Leela, but in (6.38a) the vector verb “take” encodes the agent as the main beneficiary of the building event. In (6.38b), the vector verb “give” encodes someone else as the main beneficiary of the house, since Leela is building it for someone else. A similar analysis can be given for many examples of this construction type with “give” across the data analyzed here. Consider the Kotia Oriya and Gujarati examples in (6.37a-b) above. Example (6.37a), from Kotia Oriya, is almost identical to (6.37b), where the builder of the house is not the one who is going to actually live in the house.

(6.38a) *leela=ne makān banā li-yā*
Leela=ERG house make take-MSG
‘Leela made a house (for herself).’ (Hindi, Poornima 2012:203)

(6.38b) *leela=ne makān banā di-yā*
Leela=ERG house make give-MSG
‘Leela made a house (for someone else).’ (Hindi, Poornima 2012:203)

Poornima 2012, building on the benefactive analysis, argues that it does not neatly fit with attested examples of “give” as a vector verb in Hindi. She shows that non-self affectedness and maximal change in the affected participant better describe the semantics

associated with "give" in this construction, and the benefactive function of "give" in this construction, while attested, is only secondary (212-215).

Poornima argues that in Hindi, the beneficiary of the event expressed by clauses with "give" vector verbs is often identical to the P argument, or is someone associated with the P argument in a part-whole relationship, and thus is not a semantic "benefactive." Consider example (6.39a) below (Poornima 2012, citing Nespital 1997), where the person whose lips are colored can only be Rajni, and not the A argument, Rajni's friend (as indicated by the subscript indices in the translation of this example). The A argument in (6.39a) is not directly affected by the events described in the clause.

(6.39a) *rajnī=kī sahēlī=ne apnē hāt^{h}=se us=ke hõt^{h} rãng di-ye*
Rajni=GEN.F friend=ERG REFL hand=COM 3SG=GEN lips color give-MPL
'Rajni$_{(i)}$'s friend$_{(j)}$ colored her$_{(i)}$ lips with her$_{(j)}$ own hands' (Hindi, Poornima 2012:212)

Similar examples are found throughout the data. For example, in (6.39b) from Kotia Oriya, the person who was suffering from frost is expressed in the clause by the locative post-positional phrase *tar upre* 'upon her'. The frost itself is not directly affected here. In (6.39c) from Gujarati, the person becoming well versed in science is the R argument of the entire clause, which is coded like a P argument and is flagged by the Gujarati Primary Object marker. Example (6.37b) above, from Gujarati as well, also illustrates this point.

(6.39b) *ta-r* *upr-e* *kakor rokoi* *de-la*
DEM-GEN up-LOC frost pour.on.CV give-PST.3SG
'Frost fell on her.' (Kotia Oriya, Gustafsson 1973a:168)

(6.39c) *hû* *tə=ne* *bed^{h}-ã* *shashtr-o-mã* *parõgət* *kər-i* *də-ish*
1SG 2SG=OBJ all-NPL science-PL-LOC well.versed do-CV give-FUT.1SG
'I will make you well versed in all the sciences.' (Gujarati, Christian 1987:282)

The goal of this subsection has been to establish the use of "give" in the V-V compound verb auxiliary construction type across Indo-Aryan. Further, it was shown that the function of "give" in this construction type shares at least some similarities across the family, and that it is used when the agent of the action is not an affected party of the event described by the clause. In order to exactly pinpoint the slight differences in usage of "give" in this construction type across the family, more analysis and more data are necessary.

CHAPTER VII

TOWARDS A DIACHRONY OF "DO/MAKE" N-V COMPLEX PREDICATION

One of the hallmark of Indo-Iranian languages, as well as one of the most common patterns found in the data with "do/make", is N-V complex predication (Noun-Verb complex predication) constructions. In these constructions, the structure and semantic contents of the clause are jointly motivated by a noun, encoding the situation expressed in the clause, and the verb, which encodes grammatical categories. The clause in (7.1), for example, expresses not the creation of a telephone, but rather the prototypical action of using one. Indeed, as a number of authors has noted (e.g., Telegdy 1953), the number of distinct verbal lexemes used in natural discourse by speakers of some Indo-Iranian languages, especially New Persian, is reduced, and lexical verbs are being replaced by a growing number of N-V complex predicates expressing equivalent content.

(7.1) *man telefon kard-am*
1SG telephone do.PST-1SG
'I called' (New Persian, constructed example)

This chapter has two goals. The first goal is to compare the grammar of N-V complex predication involving "do/make" across Indo-Iranian. This comparison concentrates on coding properties when "do/make" is accompanied by a deed NP element encoding the situation expressed by the entire clause, plus two other elements, encoding the doer and the affected participant.

Such construction types are the data for the second goal of this chapter, which is to propose a set of diachronic pathways of morphosyntactic change from lexical or

"heavy" uses of "do/make", to the construction types observed in N-V complex predication across the family. I will argue that complex predication in Indo-Iranian is not a single phenomenon diachronically, and that several possible pathways lead to the observed patterns. I follow particularly Haig 2002, Mohanan 1994, 1997, and Liljegren 2010 who argue in favor of a synchronically non-unitary analysis of N-V complex predication in Kurmanji, standard Hindi, and Palula. The pathways offered below, then, end up giving rise to these distinct types of N-V complex predication.

Section 7.1 gives a brief overview of the different characterizations of complex predication found across the Indo-Iranian literature, focusing on a functional, rather then structural characterizations. This is followed in Section 7.2 by presentation of a taxonomy of N-V complex predicate constructions in Indo-Iranian based on the coding properties of different components of these constructions. This taxonomy follows those presented for Hindi (Mohanan 1994, 1997), Kurmanji (Haig 2002), and Palula (Liljegren 2010). Section 7.3 argues that N-V complex predication in Indo-Iranian is potentially the result of several distinct evolutionary pathways. Section 7.4 shows that the pathways identified in 7.3 converge on a small number of distinct types, presented in 7.2. Finally, this chapter ends with a brief summary.

7.1 Delimiting "Complex Predication"

The terms "complex predicate" and "complex predication" have been used to refer to quite a number of distinct phenomena crosslinguistically. These phenomena sometimes partially or fully overlap, but often they are completely disjoint. For example, in the influential volume edited by Alsina et al. (1997), which is dedicated to complex

predication, the term is used in several distinct ways by different authors: Durie uses it to refer to a set of serial verb constructions, Rosen uses it to refer to some auxiliary constructions, especially in Romance languages, and Alsina uses it to refer to morphologically complex verb forms which include valence-changing morphology, with a focus on Bantu.

Two further papers in that volume apply the term "complex predicate" to two distinct types of constructions in Hindi and Urdu, two closely related Indo-Aryan languages. Butt applies the term to constructions composed of two verbs, as illustrated in (7.2a) below. This construction type has usually been referred to in the Indo-Aryan linguistic literature as a "compound verb" (e.g., Hook 1975, 1991, Masica 1991:326; see also Butt 1995, 1997 and Chapter VI above). In the same volume, Mohanan uses the term "complex predicate" to refer to combinations of a noun or adjective and a verb in which the lexical situation expressed by the clause is encoded by the noun or the adjective, and the structure of the clause is determined jointly by the noun/adjective and the verb, to varying degrees (see below and Mohanan 1994, 1997). This type of complex predicate is shown below in (7.2b) below.

In the Verb-Verb compound verb construction in (7.2a), the main lexical verb *banā* 'make, create' is followed by *liyā* 'take' which usually entails a completive / perfective aspect and expresses that an action that was done for the benefit of its instigator. In the Hindi N-V complex predicate in (7.2b), the main lexical verb is "do/make" and the nominal component, expressing the propositional content of the clause, is *bʰarosā* 'reliance, trust'. This is the phenomenon that is referred to here as "N-V complex predication."

(7.2a) *anjum=ne hār banā liyā*
Anjum.F=ERG necklace.M make take.PRF.MSG
'Anjum made the necklace completely' (Urdu, Butt 1997:132; glosses slightly adjusted)

(7.2b) *rām=ne mohan=par bʰarosā kiyā*
Ram=ERG Mohan=on reliance-N do.PRF.NSG
'Ram relied on Mohan' (Hindi, Mohanan 1997:434; glosses slightly adjusted)

The varied nature of the phenomena to which the terms "complex predicate" and "complex predication" have been applied has been noted by many scholars interested in the specific manifestations of this phenomenon in Indo-Iranian (e.g., Haig 2002, Samvelian 2012:18). In general, then, it seems that scholars have used these terms to characterize constructions in which the predicate of the clause is composed of two or more elements, each contributing to the overall semantics and structure of the clause.[67] In this chapter, I concentrate on what Mohanan 1994, 1997, as well as Haig 2002 and many others, have termed "N-V complex predicates." These are constructions in which the predicative weight of the clause is shared by a nominal element and a verbal element: mono-clausal constructions in which a noun and a verb "jointly determine the structure of the clause" (Mohanan 1997:432).

[67] The question of whether this contribution is compositional or holistic in nature is a separate, often debated, issue and will be touched upon only briefly below. For now, it seems that at least some N-V complex predicates are not compositional (e.g., Liljegren 2010, Samvelian & Faghiri 2013), while others (presumably especially those in the early stages of the development of N-V complex predication) are less holistic in nature.

The problem with such a characterization, as noted by Haig (2002), is that it is rather informal, and its operationalization in different languages, and even across N-V complex predicate types in a single language, is not trivial. With this characterization, for example, all valence changing constructions, whether analytic or synthetic, are instances of complex predication. Similarly, all auxiliary constructions which are correlated with alignment splits would be instances of complex predication: the main lexical verb determines the number and semantic role of participants, while the auxiliary verb determines their coding properties.[68]

Applying this characterization to N-V complex predication entails that testing for this status means asking if the structure of the clause is somehow jointly determined by the main lexical verb and the nominal component of the clause. In N-V complex predicates, the nominal component is used to encode the type of situation expressed by the clause, and interacts, at least potentially, with valence and/or argument structure. That is, the nominal component could have a role in determining the number and flagging of participants overtly expressed in the clause.

This is illustrated by example (7.2b) above. Analyzing this example, Mohanan (1997) argues that the argument structure of the clause is jointly determined by both "do/make" and *bʰarosā* 'reliance'. This includes the syntactic status of *rām* as an A argument and *bʰarosā* as a P argument, both of which are determined by the verb, and the existence and flagging of the oblique object *mohan=par* 'on Mohan', which is

[68] Acknowledging that clause structure is potentially correlated with, or is determined by, several factors (for example, verbal lexeme and type of auxiliary,) might actually be a beneficial outcome of keeping this definition in mind. This is especially true if one considers the influence of source constructions on clause alignment and valence.

determined by the nominal component of the complex predicate. It is difficult to motivate the overt expression and flagging of such oblique objects by the basic semantics of "do/make" alone, as it is not an integral element in creation or caused change-of-state events. Thus the expression and flagging of the oblique object in (7.2b) are motivated by the semantics of the deed element, which can thus be interpreted as the nominal component of a N-V complex predicate.[69]

The property of N-V complex predicates which is focused on in this chapter is the shared determination of valence and argument structure by their nominal and verbal components. This focus is not unproblematic, as it cannot identify N-V complex predicates in which the nominal component does not add to the overall valence of the clause (e.g., when the clause expresses a monovalent event or a state), and does not motivate a change in argument structure. Consider, for example, instances such as (7.3) below, from Kupia, where "do/make" is combined with *tappasu* 'ascetic'.[70] Together, *tappsu* and "do/make" denote an event of doing penance. Doing penance does not motivate the addition of another participant to the clause. Thus, even if the combination of *tappasu* and "do/make" in Kupia somehow determines together the argument structure of the entire clause, a change in valence or argument structure is not likely.

[69] Note that the syntactic status, as a P argument or not, of the nominal component of the complex predicate does not play a crucial (or any) role in the argument for a complex predicate status of (7.2b).
[70] This is originally an Indo-Aryan word, documented in both Sanskrit and Pali, where it roughly means a non-Buddhist ascetic. However the form of the word in Kupia, however, especially due to the final *u*, suggests that it was borrowed from Indo-Aryan into Telugu and then re-borrowed into Kupia.

(7.3) *tappasu ker-l-i taruwata wistnumurti otta utr-a ker-a*
penance do-PST-1SG then Wistnumurti there descend-NF do-NF
'When I've done penance, Wisnumurti will descend there.' (Kupia, Christmas & Christmas 1973a:47)

As the semantics of *tappasu* 'penance, ascetic' in (7.3) does not motivate the addition of an any argument, one cannot argue for N-V complex predicate status of the combination *tappasu* + "do/make" by using a parallel argument to Mohanan's for the Hindi combination of *bʰarosā* + "do/make." An interaction of the nominal component with valence cannot positively shown in instances where the type of situation expressed by the clause is a state or a monovalent event.

One contribution of nominal components can be at the level of event or Aktionsart semantics. For instance, events of creation, in which "do/make" is accompanied by a NP expressing the thing to be created can be considered accomplishments or, less commonly, achievements, as they are telic (following Vandler 1967). The lexical semantics of *tappasu* 'penance, ascetic' + "do/make" can be analyzed as an activity, as it is not punctual and does not have an end state.[71] Under the characterization of complex predication above, however, the structure of the clause needs to be jointly determined by the nominal and verbal components of the putative complex predicate. As the Aktionsart semantics of the event expressed in the clause does not interact with coding properties in Indo-Iranian, it is not used here to determine N-V complex predicate status of a combination of "do/make" and a noun.

[71] This is unlike situations where someone does *e:ku donni wersu tappasu* 'a two-year penance', which has accomplishment Aktionsart.

Some analyses of N-V complex predication in Indo-Iranian have identified recurring properties of the nominal component that distinguish it from prototypical NPs and P arguments. One major property is a decrease in the number of nominal categories expressed by the nominal component of the complex predicate, and a decrease in the possibility of modifying the nominal component. This has been noted for Hindi (Mohanan 1994, 1997), Kurmanji (Haig 2002), and New Persian (Müller 2010, Megerdoomian 2012, Samvelian 2012, *inter alia*). This reduction in prototypical nominal properties is sometimes accompanied by changes in stress patterns and other properties (see survey in Müller 2010), which motivated Goldberg (2003) to argue that the nominal component of N-V complex predicates is incorporated into the verb.

However, that the nominal component of N-V complex predicates in Indo-Iranian is not morphologically or syntactically incorporated into the verb, but is an independent, albeit reduced or stripped, phrasal constituent (Samvelian 2012, Müller 2010). Evidence for the phrasal status of the nominal component of New Persian Complex Predicates includes the possible, though uncommon, ellipsis of the verbal component of N-V complex predicates, as in (7.4a), and the separability of the nominal and verbal component by the New Persian Future Auxiliary construction, as in (7.4b). In (7.4a), from Müller 2010 (who cites Megerdoomian 2002) *kardan* 'do/make' is not overtly mentioned (i.e., it is "gapped", in Müller's terminology) following the coordinator *va* 'and'; but a N-V Complex Predicate *ehsas kardan* 'feel (lit. feeling make)' is still argued by Megerdoomian and Müller to be present in an underlying structure. Example (7.4b) demonstrates that in a New Persian Complex Predicate, the nominal component can be

separated from its main lexical verb by the future auxiliary. Hence the nominal is not morphologically incorporated into the verb.

(7.4a) *ta diruz ne-mi-tavanest-am* [*har.antʃe fekr mi-kard-am*
until yesterday NEG-IND-can.PRS-1SG whatever thought IND-do.PST-1SG
va ehsas ___] *boruz dah-am*
and feeling reveal give.PRS-1SG
'Until yesterday, I could not reveal what I thought or felt.' (New Persian, Müller 2010:612; glosses and transcription slightly altered)

(7.4b) *(man) telefon xah-am kard*
(1SG) telephone want.PRS-1SG do.PST
'I will telephone.' (New Persian, Müller 2010:614; glosses and transcription slightly altered)

These tests show that at least in some languages,[72] the nominal component of the complex predicate is not morphologically incorporated into the verb. The status of the nominal components of N-V complex predicates across Indo-Iranian is reminiscent of "noun stripping" (Miner 1986, Mithun 1984), which has been reported primarily for North-American and Oceanic languages. In noun stripping, a noun, which often denotes an object argument, is "stripped" of many properties associated with prototypical nouns in that specific language, but is not morphologically incorporated into the verb. Consider Kupia *tappasu* 'penance, ascetic', for example: when it encodes the situation expressed

[72] In languages where all auxiliaries follow the main lexical verb, the "future tense" test in (7.4b) cannot be used. In languages which do not allow verbal ellipsis, tests like (7.4a) are also meaningless. Furthermore, the test in (7.4a) is based on an extremely rare syntactic construction, whose absence from corpora cannot be used to argue against its availability to speakers. Finally, while Müller (2010), following Megerdoomian (2002), considers (7.4a) acceptable and grammatical, not all Persian speakers agree.

by the clause as in (7.3) above, it is never pluralized nor flagged by the object marker *=ka*, which marks identifiable, referential P arguments and all R arguments. *Tappasu* can, however, be flagged by *=ka* when it is the P argument of some other main lexical verb, such as *mec* 'accept' in (7.5):

(7.5) *tuːwi ker-l-i tappasu=ka aːŋwu mec-upa ja-l-ay*
2SG do-PST-F penance=OBJ 1SG accept-PTCP go-PST-1SG
'I have accepted the penance that you did.' (Kupia, Christmas & Christmas 1973:48)

Most other behavioral tests for the syntactic status of the nominal component of putative N-V complex predicates also involve language-specific constructions. Samvelian (2012) notes that in N-V Complex Predicates in New Persian, the nominal and the verbal components are used together in an answer to the question "What have you done?" If the nominal component behaves like a P argument, then the answer can be composed that nominal element alone, without attaching it to "do/make" or to other verbs used in complex predication.

In the Eastern Indo-Aryan, Odian languages, Kotia Oriya and Kupia, the "do/make" auxiliary construction is used in a tail-head linkage construction. In this construction, speakers often repeat the main lexical verb of the preceding clause without any of its arguments, followed by the "do/make" auxiliary,[73] as illustrated in (7.6). *Ponos sop* 'jackfruit sap,' the P argument of "bring," is not repeated in (b). In (7.7), on the other

[73] This auxiliary construction is used to express two events as thematically related, yet distinct and non-overlapping; see Chapter IV for analysis.

hand, the nominal component of the N-V complex predicate in (a) is repeated in (b). Thus the nominal component in Kotia Oriya N-V Complex Predicates is different from a prototypical P argument.

(7.6) (a) *ponos sop an-bar*
jackfruit sap bring-INF
(b) *an-i kor-i ʈeɖe-bol kodli goc mar-bar ...*
bring-NF do-NF DEM-time banana tree hit-INF …
'(a) We will bring the sticky sap of a jackfruit; (b) having brought it, we'll cut a banana tree' (Kotia Oriya, Gustafsson 1973a:150)

(7.7) (a) *ʈeɖe-bol ɖokr-i bicar ko-la*
DEM-time old.woman-F thought do-PST.3SG
(b) *bicar kor-i sob ɖokr-i tiar kor-i*
thought do-NF sap old.person-F ready do-NF
baʈ-e ne-i songoi de-la
path-LOC take-NF put.down give-PST.3SG
'(a) In this way the old woman made her plans; (b) having made plans, she prepared the sticky woman (figure) and put it in the path' (Kotia Oriya, Gustafsson 1973a:150)

Another behavioral property cited in relation to N-V complex predication is the general incompatibility of N-V complex predicates with promotional passives (e.g., Haig 2002). Specifically, it has been argued that the nominal component of N-V complex predicates cannot be promoted to the S argument of a passive construction. However, Mohanan (1994:228-229) shows that Hindi N-V complex predicates are at least sometimes used in the Passive, and thus this test does not hold for Hindi.

These behavioral tests do not directly test for the complex predicate status of a given N-V combination, as they do not directly interact with the definition of N-V complex predicates, namely joint determination of the structure of the clause and the situation expressed. Instead, they show that the nominal component of some N-V complex predication constructions does not behave like a prototypical P argument (Haig 2002 interprets the passive-test this way). The focus of the following section of this chapter is the type(s) of N-V complex predicates in which joint determination of valence and argument structure can be justified.

7.2 Toward a Taxonomy of Coding Properties of N-V Complex Predication Construction Types

The last section showed that argument structure and valence, used by Mohanan (1997) as a test of complex predicate status, might unintuitively exclude N-V complex predicates in which the nominal component does not motivate changes in argument structure or valence. But one could apply the behavioral tests mentioned in the previous section to investigate whether a specific combination of a given noun and verb behaves like known N-V complex predicates.

Using such behavioral tests, however, entails that the operationalization of the definition of N-V complex predication shifts away from the one provided above to one based also on other criteria, such as reduced behavioral similarity between the nominal component of complex predicates and prototypical P arguments. Such a shift in the operationalization of N-V complex predication is warranted in language-specific studies.

This shift, however, is less warranted in comparative studies because some of the behavioral tests, such as passivization, are based on language-specific properties of N-V complex predicates and cannot be easily applicable crosslinguistically. Thus, English (e.g., *make a claim*) and (some) Hindi N-V complex predicates can be used in passive clauses, and if passivization is regarded as a defining characteristic of N-V complex predication, these would be left out. This, in turn, would lead to an artificial decrease in the identification of crosslinguistic and intralinguistic variation in N-V complex predication constructions.

This chapter focuses, then, on N-V structures where valence and argument structure are jointly motivated by the verbal and nominal components in the predicate. The nominal component of the N-V complex may motivate the additional mention of an affected participant and its flagging. The affected participant in N-V complex predicates with "do/make" is found across the data coded as a P argument, an oblique object, an adnominal modifier, or a complement clause. Examples for this coding were amply given in Chapter IV above, and are just briefly repeated here.

In (7.8) – (7.9), from Sivandi, the affected participant is coded as a P argument. In both, the nominal component of the N-V complex predicate is *ɣond* 'collection, collect', and the affected participant is coded as a P argument. In (7.8a) *pīla* 'money' is an unflagged NP, as it is non-specific and non-identifiable in the context, and in (7.8b), the affected participant, *monajjemga* 'astrologers', is flagged by *=rā,* the Sivandi Differential Object marker as it is specific and identifiable in its context.

(7.8a) *hamī pīla γond=mā kerd*
1PL money collect=1PL do.PST
'We collected money.' (Sivandi, Lecoq 1979:93)

(7.8b) *monajjemga=rā γond=eš kerd*
astrologer.PL=OBJ collect=3SG do.PST
'He assembled the astrologers.' (Sivandi, Lecoq 1979:137)

The range of event types expressed by N-V complex predicates with the affected participant coded as a P argument is broad. The clauses in (7.9a-b) express events which are relatively high on the transitivity scale, involving metaphoric and literal ruin.

(7.9a) *hōrī ghar barabād kar dē-g-ā*
Hori house ruin do give-FUT-MSG
'Hori will ruin (his) house' (Hindi, Premchand 2017[1936]: 4.33)

(7.9b) *u=š kodak tabāh kard*
and=3SG child ruin do.PST.3SG
'she destroyed her child' (Middle Persian, AWN 64.5)

This type of N-V complex predicate also expresses events lower on the transitivity scale such as the expression of emotions or perception/cognition events, as in (7.10a-b). The clause in (7.10a) involves an event of remembering, and the remembered participant is expressed by first person Object form *mrʔ.* In (7.10b), from Kotia Oriya, the contents of the thought are expressed by *se-ʈa,* a demonstrative pronoun. Other instances of *bicar* + "do/make" are accompanied by a complement clause.

(7.10a) *w=ʔyʔd kwn-nd rstg-ʔn šwmʔ mrʔ*
and=memory do.PRS-3PL refugee-PL 2PL 1SG.OBJ
'and your refugees will remember me…' (Early Judeo Persian, Gindin 2008, 22:25)

(7.10b) *se-ʈa goʈek pila bicar ko-la*
DEM-CLF INDEF child thought do-PST.3SG
'this was what the boy was thinking' (Kotia Oriya, Gustafsson 1973a:153)

The affected participant in many N-V complex predicates across Indo-Iranian may also be expressed as an oblique object or a complement clause, as shown in (7.11a-c). In (7.11a) from Hindi, the affected participant is flagged by the comitative marker; in (7.11b) from Sivandi, it is flagged by the preposition *berey* 'on, upon'; in (7.11c) from Kotia Oriya, is it flagged by the locative marker *–e*.

(7.11a) *ab tum=se bahs kaun kar-e bhai!*
now 2SG=COM dispute who do-3SG Brother!
'now, who will argue with you, brother!' (Hindi, Premchand 2017[1936]: 3.22)

(7.11b) *ey vazīr, fekr=īt kerd berey farmā me*
hey officer, thought=2SG do.PST on dream 1SG.GEN
'hey officer, have you thought about my dream (or not)?' (Sivandi, Lecoq 1979:137)

(7.11c) *bin des-e basa ko-lai*
other land-LOC movement do-PST.3PL
'they went to another land' (Kotia Oriya, Gustafsson 1973a:201)

The status of the deed element in N-V complex predicates is not straight forward, especially in those constructions in which the affected participant is not coded as a P argument. The deed elements in (7.11a-c) are expressed by morphologically unflagged NPs, and there are no other clear candidates for P argument status in these clauses. In fact, the only constituents which coding properties do not directly contradict those of P arguments in (7.11a-c) are the NPs encoding the type of situation expressed in the clause. Thus, one might wish to argue that in (7.11a-c), these NPs are P arguments.

These arguments, when applied to these and many other examples found across the data, amounts to arguing that the deed element is a P argument from default, as when it comes to overt coding properties of P arguments, these deed elements often exhibit none. In many Indo-Iranian languages, including Sivandi and Kotia Oriya, P arguments are optionally flagged by DOM (*=rā* in Sivandi and *=ke* in Kotia Oriya), but this flagging is not found in (7.11a-c) or in any other similar clause. Such flagging of the deed element is rare to non-existent in N-V complex predicates across Indo-Iranian.[74]

The only overt coding property of P arguments which is identified with the nominal component in N-V complex predicates is verbal indexing in ergative constructions. This is illustrated in (7.12) from Hindi, and is also found in Middle Persian, Kurmanji, and other languages with constructions in which the verb indexes an absolutive argument. In (7.12) the verb indexes a feminine singular argument. The only appropriate element in the clause is *nafrat* 'hatred', a feminine singular NP. In languages

[74] For the most part, DOM in Indo-Iranian is motivated by high referentiallity and identifiability. Elements expressing the type of event are almost by definition less likely to be highly identifiable and highly referential. Thus the absence of overt object flagging is not surprising.

without such ergative constructions, there would be no positive evidence for the coding of deed elements as P arguments.

(7.12) *ilaa=ne mohan=se nafrat kii*
Ilaa=ERG Mohan=COM hatred do.PRF.FSG
'Ilaa hated Mohan.' (Hindi, Mohanan 1997:454, glosses mine)

When the affected participant is expressed by an oblique object within a complex predicate with "do/make", then, the status of the deed element shows variation across Indo-Iranian. In some languages, like Hindi, overt coding properties show that the deed element is coded as a P argument. In some other languages, there are no overt coding properties to connect the deed element with a P argument.

Finally, the affected participant can also be expressed as an adnominal modifier of the deed element. This is shown in (7.13a-b). In (7.13a), from Hindi, the deed element *kʰeti* 'farming' is modified by "two plough(ed fields)" and in (7.13b), from Nagamese, *phaisla* 'judgment' is modified by the NP expressing the target of the judgment.[75]

(7.13a) *jʰigur do hal ki kʰeti kart-ā hai*
Jhigur two plough GEN farming do.PTCP-MSG be.PRS.3SG
'Jhigur farmed two fields' (Hindi, Premchand 2017[1936]: 4.45)

[75] Mohanan 1994, 1997 and Liljegren 2010 consider examples such as (7.13a) to be instances of N-V complex predication. It is questionable, however, whether *kʰeti* 'farming' and *phaisla* 'judgment' contribute to the structure of the entire clause, as the addition of the affected participant is NP-internal in (7.13a-b).

(7.13b) *manu-khan laga phaisla na-kor-ibi,*
person-PL GEN judgment NEG-do-IMPR
'Don't judge people.' (Nagamese, Luke 6:37)

The examples so far show that, across Indo-Iranian, N-V complex predicate constructions with an overt mention of an affected participant differ with respect to two sets of coding properties. The first set includes the degree to which the nominal component of the complex predicate is overtly coded as a P argument. The second has to do with the coding of the affected participant as a P argument, an oblique object, complement clause, or an adnominal modifier of the deed element. When taking into account the interaction of these sets of coding properties, the different N-V complex predicate constructions in Indo-Iranian can be described as belonging to five types, summarized in Table 7.1.

Table 7.1: Five types of N-V complex predicates

	Type 1	***Type 2***	***Type 3***	***Type 4***	***Type 5***
Verb indexes deed?	YES		NO		NO
Deed flagging	NONE				
Affected flagging	Oblique object	Adnominal modifier	Oblique object	Adnominal modifier	P argument
Attested in:	Hindi, Middle Persian …		Kotia Oriya, Sivandi …		All lgs. analyzed here

In terms of the coding properties listed in Table 7.1, the nominal components of N-V complex predicate constructions are of two distinct sets. The first set corresponds to *Types 1, 2* and 5 in Table 7.1, and the second set corresponds to *Types 3* and *4* in Table 7.1. The nominal component of complex predicates in the first set is coded as a P argument, at least in terms of verbal indexing in ergative constructions. The nominal

component in complex predicates of the second set is an unflagged NP, usually in proximity to the verb. These two sets have been identified by Mohanan (1994, 1997), Haig (2002), and Liljegren (2010) for Standard Hindi, Kurmanji, and Palula, respectively. Despite using different terminology, all three essentially argue that the nominal component in *type 5* N-V complex predicates is functionally, but not morphologically,[76] incorporated into the verb (in Mohanan 1997's terms, a part of PRED).

In all Indo-Iranian languages analyzed here, as well as several others discussed in the literature, it is difficult to determine based on the semantics of the event (or any other semantic properties of the deed element, for that matter) which of the five types in Table 7.1 a specific N-V complex predicate belongs to (e.g., Mohanan 1997:466, Haig 2002:26). For example, N-V complex predicates of *type 5*, in which the affected participant is coded like a P argument, have been shown in (7.9 – 7.10) to express a wide range of event types.

Despite the apparent lack of semantic factors which would help to explain the distribution of N-V complex predicate types in expressing different types of situations, there are general patterns for their distribution. N-V complex predicates in which the affected participant is coded like a P argument express a broad range of event types, ranging from highly transitive events to events lower on the transitivity scale such as perception/cognition events. When the affected participant is coded as an oblique object, the semantic range of event types expressed by the N-V complex predicate is narrower, for the most part covering only one to three semantically related situation types.

---

[76] In all the languages examined here, deed elements of these type can be syntactically separated from the verbal component of the complex predicate by at least some element, be it a clitic element, an auxiliary verb, or even another NP.

The coding of the doer and affected participants seems to follow what Haspelmath (2015, see also Lazard 1994, 2015) calls "Transitivity Prominence": the tendency of participants in bivalent events to be coded as A and P arguments, regardless of how semantically (in)transitive the event is.[77]

In order to illustrate this, I collected and analyzed the types of events expressed by clauses in which "do/make" is accompanied by three elements (doer and affected participants, and a deed element) in twelve of the sixteen languages. Across these twelve languages, I identified forty-six different language-specific constructions, expressing distinct event types, including highly transitive events like "destroy" and "harm", less transitive events like "find" and "eat", expressions of attitude towards an entity such as "like", "welcome", caused change of state, perception/cognition events, and transfer events. To these, I added "general" types of events, i.e., clauses in which the deed element is not overtly expressed or is expressed by a demonstrative pronoun or another semantically light element. While semantically "light" elements are not technically a nominal component of a N-V complex predicate, such clause constructions are included here for the sake of completeness. As a result, the forty-six language specific constructions include all language specific constructions where "do/make" is accompanied by three nominal elements, one of which is a deed element, in the twelve relevant languages.

I constructed a matrix with columns representing the different types of situations and rows representing the forty-six language specific constructions in which "do/make"

---

[77] As a reminder, in this dissertation I define A and P arguments following Comrie 1989, Lazard 1994, Payne 1997, as the most agent-like and most patient-like participants in a prototypical transitive event, such as Lazard's example "kill".

is accompanied by doer and affected participants and a nominal deed element. If a specific construction was found in the data expressing a specific type of event, then the appropriate cell in the matrix equals 1; otherwise it equals 0. This matrix is visualized in Figure 7.1 below. In this matrix each construction is given a "handle" composed of a code for the language[78] and the type of coding of the affected participant.[79] The matrix is visualized using the gplots R package (Warns et al. 2015). A red cell represents a cell that equals 1; an empty cell zero.

The rows in Figure 7.1 are sorted based on similarity: constructions which express semantically similar types of events are placed more closely together. The top rows represent constructions in which the affected participant is coded as a P argument, and the Hindi construction in which the affected participant is expressed as an adnominal modifier. The distribution of the red color in these columns shows that these constructions are used to express a wide array of situation types, including highly transitive events, but also events lower on the transitivity scale.

Other clusters of red are composed of constructions in which the affected participant is expressed by an oblique object or a complement clause. For example, constructions where "do/make" is accompanied by a complement clause tend to express PCU event like "think" or "remember" and Aktionsart type predicates like "begin" or "end", and are clustered together directly below the top cluster of constructions where the

[78] BACTRIAN Bactrian, DARAI Darai, EJP Early Judeo-Persian, GOR Gorani, HIN Hindi, KOTIA Kotia Oriya, KUPIA Kupia, MAZ Ziyarat Mazandarani, MP Middle Persian, NAG Nagamese, SIV Sivandi, SOG Sogdian.

[79] P P argument, LOC general locative marker, COMP complement clause, COM comitative, AS essive, POST post-verbal NP, GEN adnominal affected participant; otherwise, the code expresses the adposition used (e.g., PAD for MP *pad*, NIMITE for Nagamese *nimite*).

affected participant is coded like a P argument. In a similar fashion, constructions where "do/make" is accompanied by a comitative marker tend to express reciprocal events such as "fight" or "argue" and also tend to cluster together.

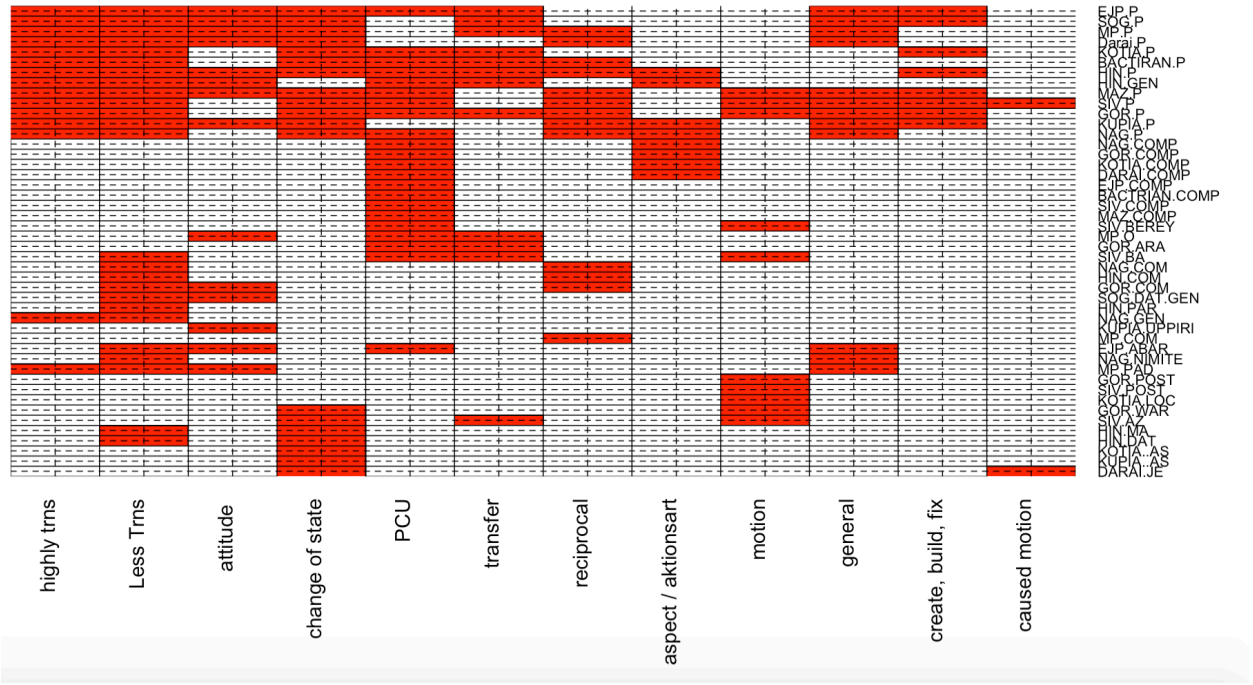


Figure 7.1 The use of 46 "do/make" constructions to express different situation types

Figure 7.2 presents a Neighbor-Net analysis (Bryant & Moulton 2004) of the matrix in Figure 7.1. In this analysis, the hamming distance between each pair of constructions, based on their use in clauses expressing specific situation types, is calculated, and a distance-based (i.e., non-phylogenetic) network is constructed based on these distances. The Neighbor-Net in Figure 7.2 was constructed using SplitsTree 4 (Hudson & Bryant 2006), a free open-source software implementing several network constructing methods. In this Neighbor-Net, one finds roughly the same clusters as in Figure 7.1. The biggest and clearest cluster is composed of the constructions in which the affected participant is coded as a P argument, together with the Hindi construction where

the affected participant is coded as an adnominal modifier. The nodes in this cluster are marked by a solid black line. The cluster of constructions in which "do/make" is accompanied by a complement clause is marked by a dotted black line. The constructions in which the affected participant is flagged by a comitative are marked by a dashed line.

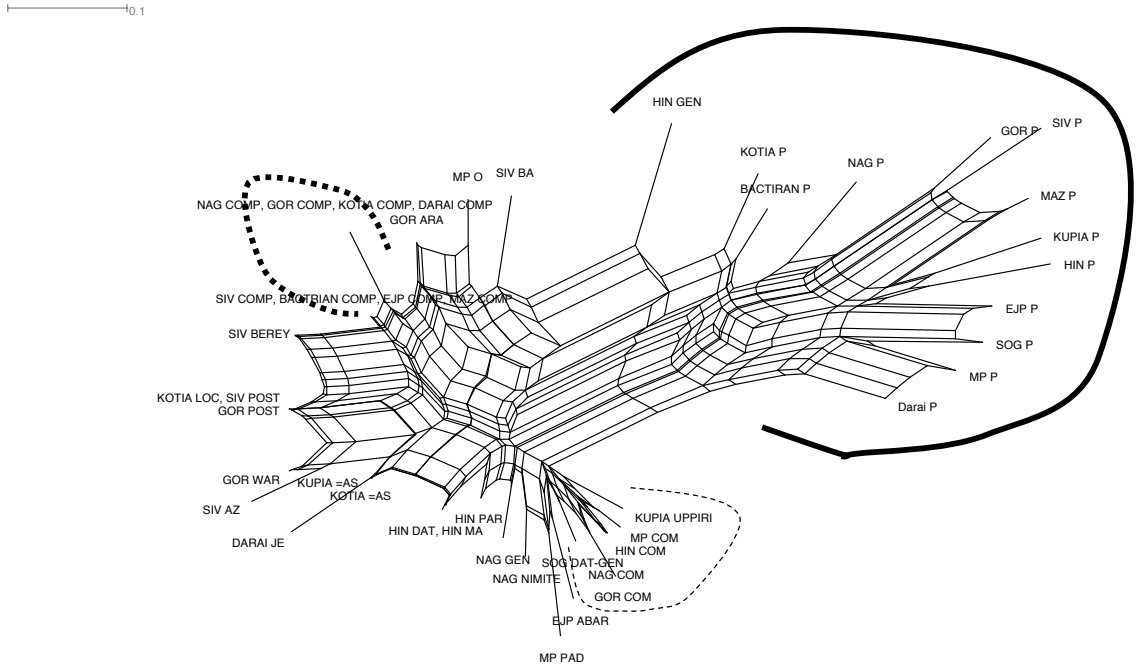


Figure 7.2 Neighbor-Net of 46 "do/make" constructions with overt affected participant

Figures 7.1 and 7.2 show that while it is difficult to predict the form of a specific N-V complex predicate based on its semantics, there are overall patterns in their distribution. N-V complex predicates in which the affected participant is expressed as a P argument express a variety of distinct event types. Those in which the affected participant is coded as an oblique object express a narrower range of event types, usually related to other uses of the specific oblique object involved. This pattern is similar to what Haspelmath (2015) and Lazard (1994, 2015) find with lexically heavy verbs: the simple

transitive argument frame expresses a wide variety of events while other clause types with two arguments express a semantically narrower variety of event.

7.3 A Non-Unified Proposal for the Evolution of N-V Complex Predicates

This section proposes that the evolution of the different types of N-V complex predicate constructions in Indo-Iranian is the result of several parallel converging diachronic processes. Further, the reason behind Haspelmath's "transitivity prominence" (2015) pattern applying also to N-V complex predicates is often the retention of the patterns of coding properties attested in previous stages in the pathway. The illustration of the main pathways argued for here are based primarily on Middle Persian, since it is the oldest language in the data in which I have found clear instances of all the pathways that are presented in this section. Another reason for using Middle Persian as evidence is that it seems, as Korn (2013) notes, that the question of the existence of N-V complex predication in Middle Persian has not yet been satisfactorily addressed.

The pathways proposed here all have "literal" or "heavy" construction types of "do/make" as their starting point. The grammatical changes culminating in N-V complex predicate construction types are all instances of reanalysis (and actualization) and extension. Some steps in these processes involve extensions of the semantic range or type of nouns expressing a particular component of the clause. Others involve changes in the coding properties of either the nominal component of the complex predicate or the constituent expressing the affected participant.

The literature on N-V complex predicates has not been entirely silent on their evolution. Ciancaglini (2011) claims that they can already be discerned in Old Iranian,

relying mainly on evidence from Imperial Aramaic and arguing that they are calques from Old Persian. Korn (2013), however, casts doubt over whether these instances can really be considered complex predicates, citing the lack of evidence for semantic bleaching of the verb, and as a result, lack of evidence for the structure of the clause being motivated by two components simultaneously.

Haig (2002) suggests that the origin of the Kurmanji Kurdish, and perhaps the Iranian N-V complex predicates in general is an old combination of the main verbal predicate with a particle in the well-known Indo-European particle-verb construction. The position, or function, of the "particle" can also potentially be filled by an adjective and, later, by a noun as well. Thus Haig suggests the following evolutionary pathway: particle + verb → adjective + verb → noun + verb. Haig does not offer extensive argumentation in favor of this very sensible pathway, but below in subsection 7.3.2 I will argue that there is some evidence for the transition from adjective + verb → noun + verb complex predicates in some of the Middle Persian data.

7.3.1 From Bivalent "Do/Make" Constructions to N-V Complex Predication

The starting point for the evolution of the N-V complex predication is the "creation" construction type of "do/make". In this construction type, "do/make" is used together with two NPs, the first expressing the doer participant and the second expressing the thing created. For the most part, the doer is coded as an A argument and the thing created is coded as a P argument. This construction type is illustrated by (7.14), where Ohrmazd creates the world using his own wisdom. This construction type is found in all the languages analyzed in which cognates of "do/make" survived.

(7.14) *ohrmazd ī xwadāy ēn gēhān pad dānāgīh kard*
Ohrmazd LNK Lord DEM world by wisdom[80] do.PST.3SG
'Ohrmazd the Lord created this world with knowledge' (Middle Persian, DK6 311)

Korn (2013) shows that already in Old Persian, one finds clauses where "do/make" is accompanied by a NP expressing a type of action or event, rather then a tangible object created, and that consequently, "do/make" has the semantics of "perform" or "carry out", rather than "create". This is demonstrated in (7.15a) from Old Persian, and (7.15b) from Hindi.

(7.15a) *avadā hamaranam akunava*
there battle.NSG.ACC/NOM[81] do.IMPF.3PL
'There they fought the battle.' (Old Persian, Bisotun II.34; cited from Korn 2013:42)

(7.15b) *gobar=ne kaʈakʃ kiyaa*
Gobar=ERG sidelong.glance do.PRF.MSG
'Gobar sneered.' (Hindi, Premchand 2017[1936]: 3.20)

Korn mentions that it is unclear whether Old Persian clauses like (7.15a) are instances of N-V complex predication or not. There is not much evidence to suggest that *hamaranam* 'battle' expresses a type of action and not a specific instance of fighting,

[80] This gloss might be a bit misleading. The internal structure of *dānāgīh* is *dān-āg-īh* and the glosses would be "know.PRS-ADJ-ABST".
[81] Indo-European nouns of Neuter gender do not formally distinguish between Nominative and Accusative case forms.

especially as the actual battle is clear from the original context of (7.15a). What is obvious is that in (7.15a-b) "do/make" is accompanied by two NPs, and there is no change in the overall valence of the clause which is motivated by 'battle' or 'sidelong glance' when compared to the structure of (7.14). In many instances, however, the events described in clauses like (7.15a-b) include one additional participant. When speakers overtly mention these participants in the clause, then it can be argued that their mention is motivated by the type of situation encoded by the deed element, and the argument structure of these clauses is motivated by both the verb and the deed element.

The first type of such a complex predication is illustrated by the Middle Persian collocation of *zaxm* 'harm, wound' with "do/make" in (7.16a-b). The clause in (7.16a) does not describe any specific instance of harm or injury, but rather refers to a general event and cites the degree of sin accrued by a transgressor. In (7.16b), *zaxm* is conjoined with a noun representing another type of affliction (*rēš* 'sore'), and both function as the Direct Object of the clause. In both examples, *zaxm* + "do/make" expresses a type of an event, rather than a specific instance of wounding.

(7.16a) *ka zaxm kun-ēd yāt-ēw*
when wound do.PRS-3SG yat-INDEF
'When he makes a wound, the (measure of sin) is one yat.' (Middle Persian, RAF A59)

(7.16b) *ka zaxm ud rēš kun-ēd ēg=eš wināh ce*
when wound and sore do.PRS-3SG then=3SG sin what
'When he inflicts wounds and sores, what is his sin?' (Middle Persian, RAF Q58)

In these examples there is no evidence that the argument structure of the clause is motivated by any constituent other than "do/make". There is no evidence from any coding property that the status of 'wound', like that of 'battle' and 'sidelong glance' in (7.16a-b) above, is any different from the status of tangible objects created by the A argument of the "do/make" verb. There might be a difference between these types of P arguments in some behavioral properties, but the search for N-V complex predication in this chapter concentrates on the emergence of changes in coding, not behavioral properties. It should be kept in mind, though, that changes in coding properties usually follow changes in behavioral properties (e.g., Cole et al. 1980, *inter alia*).

The event of making a wound, described by the collocation *zaxm* + "do/make", involves not just one participant, but at least two: the doer inflicting the injury and the person who suffers it. In (7.16a-b) above, no affected participant is overtly expressed in the clause. In some situations, however, speakers wish or need to express the participant who suffers the injury, and when these situations arise, the affected participant is usually flagged by the preposition *pad* 'in, to, by'.

(7.17) *mard ka pad zan ī xwēš zaxm kun-ēd*
man when in woman LNK REFL wound do.PRS-3SG
'a man, when he wounds his wife' (Middle Persian, RAF Q51)

The deployment of such PPs to express affected participants is not limited to combinations of *zaxm* + "do/make"; it is found also where the deed accompanying "do/make" is not a noun, but a complex phrase composed of the demonstrative *ān* followed by a relative clause as in (7.18a) below, or a contextually recoverable zero as in

(7.18b). Such clauses usually express a type of behavior, and *pad* seems to be functionally equivalent to the English (*un)to* as in *do (un)to other what you want them do (un)to you.*

(7.18a) *ān=ez ī=š ... ne nēk sah-ēd*
DEM=FOC LNK=3SG … NEG good seem.PRS-3SG
pad kas-ān ne kun-ēd
to person-PL NEG do.PRS-3SG
'He does not do unto other people that which does seem to him not good.'
(Middle Persian, DK6 C73)

(7.18b) *u=šān harw ēwēnag pad ruwān ī xwēš kard baw-ēd*
and=3PL all case to soul LNK REFL do.PST be.PRS-3SG
'and in any case, people do it to their own souls' (Middle Persian, DK6 244)

It seems, then, that in Middle Persian, when the type of event encoded by the deed element involved a potential affected participant, i.e., when the deed element had valence of its own, this participant was potentially expressed by a PP headed by *pad* 'in, to, by'. This coding of the affected participant is a general strategy used in clauses expressing events with an affected participant, regardless of whether the P argument expressing the event is a lexical noun, a demonstrative plus a relative clause, or a contextually identifiable zero. It is clear that the deployment of this PP to express the affected participant is motivated by the type of event expressed in the clause, and thus at least when the deed element is expressed by a lexical noun, the clause is an instance of N-V complex predication.

Another strategy for expressing the affected participant, found across several of the languages analyzed here, is as an adnominal modifier of the deed element. The deed element is a NP and, as such, is potentially compatible with different types of adnominal modification. As the affected participant is a part of the valence of the deed element, it is only natural that it be expressed by adnominal modification. Examples for this strategy are found throughout Indo-Iranian. Examples (7.19a-b) show this strategy.

(7.19a) *ud ka abāz ēst-ēd zan-īh ī*
and when back stand.PRS-3SG woman-ABST LNK
ān mard ne kun-ēd
DEM man NEG do.PRS-3SG
'and if she turns away, and will not perform the duties of a wife towards that man'
(Middle Persian, RAF A16)

(7.19b) *rām=ne nīnā=kī madad kī*
Ram=ERG Nina=GEN.FSG help do.PRF.FSG
'Ram helped Nina.' (Hindi, Mohanan 1997:432)

Mohanan (1997:432) lists (7.19b) as an instance of N-V complex predication, despite the fact that it could be argued that the helped participant in (7.19b) and the husband in (7.19a) are not direct constituents of the clause, but internal constituents of the deed NP, and that the argument structure of the clause as a whole is unchanged. This is unlike the status of the PPs headed by *pad* in (7.17 – 7.18) above, which are external to the NP expressing the deed element. Whether examples such as (7.19a-b) fall under the umbrella of one or another characterization of N-V complex predication, it is clear that

the valence of the deed element plays a role in them, and motivates an overt expression of the affected participant.

The third strategy used to overtly mention an affected participant evolves through the so-called reciprocal alternation (e.g., Levin 1993:58-65). In languages which exhibit the reciprocal alternation, some events in which two participants interact as equals (i.e., semantically reciprocal events) can be construed by speakers using two distinct argument structure constructions. In the first, both participants form a single constituent, usually the A/S argument of the clause. This can be seen in (7.20a) below, from Modern Hebrew, where the male and female cats together constitute the S argument of the clause, and the verb indexes a 3PL subject. In the second argument structure construction, one of the participants is construed as the S/A argument of the clause and the other is expressed as an oblique object flagged by a comitative marker. This can be seen in (7.20b), where the male cat is coded as the subject of the clause, and is indexed on the verb, whereas the female cat is expressed as an oblique object flagged by the comitative preposition.

(7.20a) *ha-χatul* *ve=ha-χatul-a* *ravu*
DEF-cat.MSG and=DEF-cat-FSG fight.PST.3PL
'The male cat and the female cat fought.' (Modern Hebrew, constructed example)

(7.20b) *ha-χatul* *rav* *ʔim* *ha-χatul-a*
DEF-cat.MSG fight.PST.MSG with DEF-cat.FSG
'The male cat fought with the female cat.' (Modern Hebrew, constructed example)

In Modern Hebrew V-N complex predicate constructions, this alternation still holds. In (7.21a-b), the predicate is composed of a combination of the verb "go, walk"

and the plural form of "punches", *makot.* In (7.21a), both semantic participants in the fight are expressed together as the S argument of the clause, while in (7.21b) the male cat is expressed as the S argument and the female cat as an oblique object flagged by the comitative marker.

(7.21a) *ha-χatul ve=ha-χatul-a halχu makot*
DEF-cat.MSG and=DEF-cat-FSG go.PST.3PL punches
'The male cat and the female cat fought (lit. 'went punches').' (Modern Hebrew, constructed example)

(7.21b) *ha-χatul halaχ makot ʔim ha-χatul-a*
DEF-cat.MSG go.PST.3MSG punches with DEF-cat-FSG
'The male cat fought (lit. 'went punches') with the female cat.' (Modern Hebrew, constructed example)

The same alternation is found in Middle Persian constructions with "do/make," in which the clause, through the semantics of the deed element, expresses such reciprocal events. In (7.22a-b), "do/make" is preceded by *kārezār* 'battle, fight'. The two clauses differ in the way the participants in the fighting event are grammatically expressed. In (7.22a), neither side of the battle is lexically expressed in the clause, but they are indexed together on the verb as a 3PL participant. In (7.22b), in contrast, one participant, "the spirit of Dron" is expressed as the A argument and is indexed on the verb. The other participant in the fight, "those demons," is expressed as an oblique object flagged with the comitative preposition *abāg* 'with'.

(7.22a) *pad arang bār sē kārezār kun-ēnd*
on Arang side three battle do.PRS-3PL
'On the bank of the Arang they will fight three battles.' (Middle Persian, ZWY 7.12)

(7.22b) *mēnōg ī drōn ka yaz-ēnd*
spirit LNK dron when pray.PRS-3PL
abāg awēšān dēw-ān kārezār kun-ēd
with 3PL demon-PL battle do.PRS-3SG
'The Spirit of Drōn, when (people) perform the ceremony, fights against those demons.' (Middle Persian, PRDD 9)

A similar alternation can be found when the deed element is *pašn* 'contract, agreement' and its combination with "do/make" means roughly "agree". In (7.23a), the parties to the agreement jointly form the A argument of the clause and are indexed together on the verb. In (7.23b), in contrast, one of the parties to the agreement, the man, is coded as the A argument and is indexed on the verb. The other party to the agreement, the man's wife, is coded as an oblique object flagged by the same comitative preposition as in (7.22b) above.

(7.23a) *ka ō kārezār šaw-ēnd be ka pašn kun-ēnd*
when to battle go.PRS-3SG unless when contract do.PRS-3PL
'When they go to battle, unless they make an agreement.' (Middle Persian, PURS A55)

(7.23b) *mard-ēw ke abāg zan ī xwēš pašn kun-ēd*
man-INDEF REL with woman LNK REFL contract do.PRS-3SG
'a man who makes an agreement with his wife' (Middle Persian, RAF Q62)

This subsection has presented three scenarios in which "do/make" is accompanied by a P argument expressing a type of situation, and the affected participant of this situation is lexically mentioned in the clause. This affected participant is found overtly expressed by one of three strategies: a PP headed by *pad* 'at, to, by', an adnominal modifier, or a comitative phrase (usually a PP). The overt lexical expression of the affected participant is motivated by the type of event the clause expresses, encoded as the P argument. Where the affected participant is expressed as an oblique object, the structure of the clause is jointly determined by the verb and the deed element, and thus the clause is an instance of complex predication. Where the affected participant is expressed as an adnominal modifier, the situation is murky and at least in some interpretations of "structure of the clause," such clauses would not be instances of N-V complex predication.

7.3.2 From Caused Change of State to N-V Complex Predication

The second type of evolutionary pathway proposed here is composed of six successive and dependent stages. This pathway is sketched in Table 7.2. The second column in the table provides a schematic description of the different components of the construction types representing each stage, and the third column provides an English rendering (often ungrammatical in English) of the type of tokens which belong to that stage. The schematic structure depicted by the second column does not presuppose a formal model of representation of each construction type, and is possible to formalize by most constraint-based formal representations, such as those proposed by Müller (2010) or Mohanan (1994, 1997), *inter alia*.

Table 7.2: The evolution of N-V complex predicates.

Stage	Schematic structure	Schematic example
Stage (1)	$NP_{[A\ argument]}$ $NP_{[P\ argument]}$ ADJ "do/make"	She made him happy
Stage (2)	$NP_{[A\ argument]}$ $NP_{[P\ argument]}$ $NP_{[state\ semantics]}$ "do/make"	She made him (a/her) shelter
Stage (3)	$NP_{[A\ argument]}$ $NP_{[P\ argument]}$ $NP_{[transitive\ event]}$ "do/make"	She made him punishment
Stage (4)	$NP_{[A\ argument]}$ $NP_{[P\ argument]}$ $NP_{[less\ transitive\ event]}$ "do/make"	She made him advice
Stage (5)	$NP_{[A\ argument]}$ $OBL_{[oblique\ argument]}$ $NP_{[less\ transitive\ event]}$ "do/make"	She made advice to him
Stage (6)	$NP_{[A\ argument]}$ $OBL_{[oblique\ argument]}$ $NP_{[event;\ P\ argument]}$ "do/make"	She made trust on him

The initial stage in this scenario is clauses in which "do/make" is used to express a caused change of state. In this type of utterance, the doer is coded as an A argument and the affected participant whose state changes is coded as a P argument. The new state of the affected participant is expressed by a lexical adjective which is morphologically unflagged. This stage is found in many, if not all, Indo-Iranian languages in which cognates of "do/make" have been identified. This first stage is illustrated in (7.24a-b), from Middle Persian and Ziyarat Mazandarani. The status of the affected participant, i.e., the first person in (7.24a), as a P argument is clear since it is indexed on the verb. The status of the affected participant in (7.24b), the house, is apparent as it is flagged by the Ziyarati Object marker *=re*.

(7.24a) *a=t* *arzān-īg-tar* *kard* *h-ēm*
then=2SG worth-ADJ-CMP do.PST be.PRS-1SG
'you made me worthier' (Middle Persian, AWN 4.14)

(7.24b) *ta* *xunā=re* *per kārd-im*
until house=OBJ full do.PST-1PL
'(we brought firewood), until we filled our house' (Ziyarat Mazandarani, Shokri et al. 2013:119)

In Mohanan (1997), clauses like (7.24a-b) are considered to be instances of N-V complex predication in which the adjective phrase participates in determining argument structure of the entire clause. For example, Mohanan considers the Hindi adjective-verb combination *saaf kiyaa* 'clean' (lit. 'clean do') to be an instance of complex predication. This seems to stem from the fact that for Mohanan, the basic argument structure of "do/make" is the transitive one used to express creation, so that the adjectives in (7.24a-b) above or in *saaf kiyaa* contribute to a difference in argument structure. Crosslinguistically, however, as well as across Indo-Iranian, "do/make" verbs can often be considered compatible with two basic argument structures. The first argument structure involves creation and the second involves caused change of state.[82] Whether this use of "do/make" is basic or already a complex predicate, it does not change the fact that it is a diachronic source for many types of N-V complex predication.

The second stage in the pathway is characterized by the use of nouns rather than adjectives to express the state of the affected participant. These nouns can either express states directly or by implicature, or carry a state semantics by some metaphorical or metonymical extension. The coding properties of clauses of this type are identical to the

[82] In Modern Hebrew, for example, the verb *asa* 'do/make' can be used in both types of argument structure constructions: e.g., *asiti ʃiʔurim* "I did homework" vs. *haʃiʔurim asu oti ʔayef* "The homework made me tired." Similarly, English *make* can be used in both such argument structure constructions as in *I made rice* and *The paste made the rice spicy*.

coding properties of clauses of stage 1. Stage 2 clauses are illustrated by (7.25), where the new state is expressed by *panāh* ‘shelter’, a lexical noun.

(7.25) *panāh-ē ān weh ke xēm ī xwēš panāh kun-ēd*
shelter-INDEF DEM good REL character LNK REFL shelter do.PRS-3SG
‘This shelter is best, a man who makes his character (into a) shelter.’ (Middle Persian, DK6 252)

In stage 3, the range of possible lexical nouns expressing the new state expands to include nouns expressing highly transitive events that imply a change of state of the affected participant. In (7.26), the clause expresses an event of destruction. The deed, or type of event, is expressed by *tabāh* ‘ruin, destruction’ and the affected participant is expressed as a P argument. The coding properties in (7.25) are identical to those is (7.26).[83]

(7.26) *u=š kodak tabāh kard*
and=3SG child ruin do.PST.3SG
‘She destroyed her child.’ (Middle Persian, AWN 64.5)

Clauses which can be classified as stage 3 can already be analyzed as instance of complex predication. The situation expressed is encoded by the non-verbal, or nominal, element, and not by “do/make”. The affected P participant is the undergoer of the event

[83] The combination *tabāh* + “do/make” is very common in Middle Persian, and all occurrences of *tabāh* in the data available to me were in constructions such as (7.26) or as a complement of “be/become”.

expressed in the deed element. Diachronically, the coding of the affected participant as a P argument in (7.26) is a retention of the coding properties in stage 2. But as the event expressed in (7.26) is highly transitive, the coding of the affected participant as a P argument is also synchronically motivated by the fact that P arguments express affected participants in highly transitive events.

In stage 4 of the pathway, the deed element is extended to express events lower on the transitivity scale which do not necessarily imply a change of state of the affected participant. That is, the extension of the deed element from an element expressing a state to an element expressing a type of event has been completed. The coding properties of the doer, the affected participant, and the deed element remain unchanged at this stage. The doer is coded as an A argument and the affected participant is coded as a P argument. This is illustrated by (7.27a-b). In (7.27a), from Middle Persian, the deed is *ayārīh* 'help, assistance', and the doer participant, the referent of the headless relative clause, is not lexically expressed but is indexed on the verb. The affected participant, the P argument, is expressed as the unflagged 3PL pronoun. In example (7.27b), from Kotia Oriya, the doer participant is not lexically expressed in the clause, and the deed is expressed by the morphologically unflagged NP *sumorna* 'prayer'. The affected participant is expressed by a NP flagged by the Primary Object marker *=ke.* Both (7.27a-b) express events lower on the transitivity scale, where the affected participant is not directly, or physically, affected by the events described in the clause. However, the coding properties of the affected participants are identical to those of a P argument, just as in stage 3. One can view examples like (7.27a-b) as illustrating of Haspelmath's "transitivity prominence"

principle, i.e. the tendency of clauses expressing events lower on the transitivity scale to code arguments as if they were arguments of prototypical transitive events.

(7.27a) *ke awēšān ayārīh kun-ēd*
REL 3PL help do.PRS-3SG
'someone who helps people' (Middle Persian, PRDD 8J)

(7.27b) *durga=ke sumorna ko-la*
Durga=OBJ prayer do-PST.3SG
'He prayed to Durga.' (Kotia Oriya, Gustafsson 1973a:122-123)

Clauses like (7.27a-b) would be classified as lower on the transitivity scale by many different semantic and functional approaches to transitivity (e.g., Hopper & Thompson 1980, Næss 2007). Speakers' recognition of decreased transitivity can be actualized by coding of a (less-)affected participant as an oblique object, often flagged by adpositions. This is illustrated in (7.28a-b), from Middle Persian, where the affected participants are coded as oblique objects and are flagged by the preposition *ō* 'to'. Note that in (7.28b) the deed element is *ayārīh* 'help, assistance', which is also found with the affected participant coded as a P argument, as in (7.27a) above.

(7.28a) *ke pad hunar abastān abāz ō tan ī xwēš kun-ēd*
REL in skill trust back to body LNK REFL do.PRS-3SG
'he who puts trust in himself regarding skill' (Middle Persian, DK6 E9)

(7.28b) *ke pad zōr ud nērōg ī man ayārīh ō weh-ān nē kun-ē*
REL in strength and power LNK 1SG help to good-PL NEG do.PRS-2SG
'(you) who do not help good people by my strength and power' (Middle Persian, PRDD 26)

The actualization from (less-)affected participants coded as P argument in stage 4 to (less-)affected participants coded as oblique objects in stage 5 rests on the availability of an oblique objects structure, either in the same language, or in another language that the speaker often uses. That is, if speakers have been exposed to oblique objects elsewhere, and if oblique objects are used to express participants in semantically related event types, then they are "available" for speakers as a source for analogy. Across the Iranian language family, as well as in many Indo-Aryan languages, the addressee of lexical question verbs is flagged by an adposition. In Sivandi, for example, the addressee of question verbs is flagged by *ba* 'to, for', as shown in (7.29a). The addressee in (7.29b), a complex predicate with *sohāl* 'question' and "do/make", is likeways flagged by *ba* 'to, from'.

(7.29a) *ī ma'nā=rā ki ba te pors-iy=eš en?*
DEM meaning=OBJ who to 2SG ask.PRS-3SG=3SG be.PRS.3SG
'Who asked you the meaning of this?' (Sivandi, Lecoq 1979:113)

(7.29b) *Ye rū=i ba mehtar sohāl=em kerd*
one day=INDEF to groom question=1SG do.PST
'One day, I asked the groom.' (Sivandi, Lecoq 1979:113)

This actualization can also be the result of contact-induced analogy. Borrowing of argument coding strategies is not unheard of in Indo-Iranian. Stilo 2004, for example, argues that an increase in the frequency of the DOM marker *=ra* in Gazi, a Central Iranian language, among younger, more educated speakers may be influenced by New Persian. Across the family, the addressee of questions is often flagged by an adpositional marker that also flags source of motion. This can be seen in the Middle Persian clause in (7.30a), where the preposition *az* 'from' flags the addressee. The clause in (7.30b), from Ziyarat Mazandarani is a N-V complex predication with *soāl* 'question' as the nominal component, and the addressee of the question is flagged by *je* 'from'.

(7.30a) *u=š az awēšān pursīd*
and=3SG from 3PL ask.PST.3SG
'He asked them.' (Middle Persian, DK6 D5)

(7.30b) *esā šemā me=je soāl ā-kon-in*
now 2PL 1SG=from question PRV-do.PRS-2PL
'Now you ask me.' (Ziyarat, Shokri et al. 2013:149)

In another instance from Sivandi, one finds *berey* 'about, for', most likely a loan from New Persian, flagging NPs with "do/make" in two semantic types of events. The first, illustrated in (7.31a), is a motion event; the second, in (7.31b), is a PCU type event.

(7.31a) *kūš=eš ker-ime berey pātaxt*
speed=3SG do.PRS-1PL for capital
'We hurried to the capital.' (Sivandi, Lecoq 1979:99)

(7.31b)*fekr=īt* *kerd* *berey* *farmā me*

thought=2SG do.PST for dream 1SG

'Did you think about my dream?' (Sivandi, Lecoq 1979:137)

The preposition *berey* is often used in Sivandi to flag goals of motion, especially "big" destinations such as cities. It is not, however, used with heavy PCU verbs in the Sivandi data analyzed here. The lack of examples can be explained by one of two factors. First, the absence of *berey* with lexical PCU verbs might be coincidentally missing from the texts published in Lecoq 1979, even though the flagging of oblique objects of perception/cognition verbs with *berey* is possible in Sivandi. The other option is that this use of *berey* has been borrowed from Colloquial New Persian, possibly together with *fekr* 'thought', an Arabic loanword.

In summary, the shift from stage 4 to stage 5 is the first step in diachronic evolution that involves a change in coding of participants. The shift is composed of several parallel instances of actualization, each potentially culminating in a different adposition flagging the oblique object. It is the shift to stage 5, then, that gives rise to the wide array of argument frames in which "do/make" is found across the Indo-Iranian family.

The shift from stage 5 to stage 6 involves a (potential) reanalysis of the syntactic status of the deed element as a P argument. This reanalysis may be actualized only in languages where it can be supported by changes in coding means. The only coding means available for such an actualization in the data is verbal indexation of absolutive arguments. Evidence for such a reanalysis is found in Middle Persian and Hindi, as

shown by (7.32), repeated here from (7.12). In (7.32) the verb indexes the P argument *nafrat* 'hatred' in gender and number.

(7.32) *ilaa=ne mohan=se nafrat kii*
Ilaa=ERG Mohan=COM hatred do.PRF.FSG
'Ilaa hated Mohan.' (Hindi, Mohanan 1997:454, glosses mine)

To summarize, this subsection has presented a pathway from caused-change-of-state use of "do/make", where the new state is expressed by a lexical adjective, to a set of several distinct N-V complex predication types. The emergence of these complex predication types depends on the availability of particular coding means as sources of analogy for actualization of different reanalyses. It seems that if speakers do not have oblique objects available to them, they are less likely to innovate them from scratch in complex predicate constructions.

7.4 Complex Predicate Conspiracy

The processes presented in Subsection 7.3 all converge upon a structurally limited range of N-V complex predicate construction types. These construction types are shown in Table 7.1 above. The present section argues that some of these construction types, especially *Type 1* and *Type 3* N-V complex predicates, in which the affected participant is coded as an oblique object, arise through several distinct processes. Thus, it seems as if the diachronic processes behind the rise of N-V complex predicates converge on a specific configuration of coding means, or a specific set of construction types.

In Middle Persian, N-V complex predication *Type 1* can be the result of three distinct diachronic pathways. The first two pathways, introduced in subsection 7.3.1, have the "do/make" creation construction as their starting point. In these constructions, the deed element starts out being coded as the P argument, and when it encodes events which have an affected participant, these participants can be lexically mentioned in the clause by a PP headed by *pad* 'at, to, by'. When the deed element encodes a reciprocal event in which a group of participants interacts as equals, then one participant can be expressed by a comitative PP headed by *abāg* 'with'. In Middle Persian, those are two distinct pathways which result in *Type 1* N-V complex predicates: the nominal component of the complex predicate is coded as a P argument through indexation on the verb in ergative clause constructions, and the affected participant is expressed as an oblique object.

Type 1 N-V complex predication also arises in Middle Persian through the pathway introduced in 7.3.2. The source for this is the caused-change-of-state construction. In this pathway, the affected participant is flagged by an adposition in stage 5, based on analogy to other oblique objects. This was shown using Middle Persian PPs headed by *ō* 'to' and *az* 'from'. This stage may be immediately followed by a reanalysis of the status of the "deed" nominal component of the complex predicate as a P argument. This reanalysis is motivated by the lack of any other clear candidate for P argumenthood, as there are no other morphologically unflagged NPs apart from the A argument.

In languages such as Middle Persian, Hindi, and Kurmanji, reanalyses of the "deed" nominal component of the complex predicate as a P argument can be actualized by its indexation on the verb in ergative clause constructions in stage 6 of the pathway

(see subsetion 7.3.2). The result is a *Type 1* N-V complex predication. Thus, in Middle Persian, as well as Hindi and Kurmanji, it seems that several pathways conspire to converge onto the same *Type 1* complex predication.

In languages such as Nagamese, Kotia Oriya, and Sivandi there are no ergative clause constructions, and hence no opportunity for the deed element to be indexed on the verb, the same convergence occurs, only this time the pathways converge on a *Type 3* N-V complex predicate rather than *Type 1*.

7.5 Summary

This chapter has argued that N-V complex predicates are neither a synchronically nor a diachronically unitary phenomenon. The fact that in some languages one can find several distinct types of N-V complex predications has already been noted by some authors, most notably Mohanan (1994, 1997) for Hindi and Haig (2002) for Kurmanji. Across the family, the types of complex predication with "do/make" differ in the coding of both the nominal component of the N-V complex predicate (i.e., whether it is coded as a P argument or not) and in the coding of the affected participant as a P argument, oblique object, or adnominal modifier. This chapter has shown that similar variation is found in languages such as Kotia Oriya and Sivandi, where the nominal component of the complex predicate is never overtly coded as a P argument.

This chapter has also proposed several pathways which lead to the joint determination of the argument structure of "do/make" complex predicates by the nominal and verbal components. Some of these pathways seem to conspire to converge onto two specific types of complex predication. These pathways have distinct source constructions

(“do/make” expressing creation, and “do/make” expressing caused change of state), and distinct processes of reanalysis and actualization.

CHAPTER VIII

THE CO-EXPRESSION OF NOMINAL PREDICATION DOMAINS IN INDO-IRANIAN

The vast majority of the "be/become" construction types discussed in Chapter V expresses one of the six nominal predication domains: equation, predicate property, proper inclusion, predicative possession, the predicate locative, and the existential. A quick glace at the examples and analyses presented in Chapter V shows that some of these domains are often expressed by identical configurations of structural coding means, while other domains are often expressed by some unique structural features and are less frequently co-expressed.[84] For example, "be/become" verbs are often accompanied by two unflagged NPs, and for the most part such clauses express equation, predicate property, and proper inclusion, as shown in (8.1a-c). In this study, I refer to these as the "core" domains of nominal predication.

(8.1a) *īrij kuř=aš biya*
irij son=3SG be.PST.3SG
'Irij was his son' (Gorani, Mahmoudveysi et al. 2012, 5:8)

(8.1b) *ma na ţing bhí-l-u*
1SG.NOM NEG firm be-PST-MSG
'(I said) I was not firm' (Palula, Liljegren & Haider 2015:52-53)

[84] I use the term "co-expression" following Hartmann et al. 2014 (see especially footnote 5). Essentially, two functional domains are co-expressed if the coding means expressing them are isomorphic, or identical. An alternative term for "co-expression", then, could be "isomorphic expression" or "iso-expression".

(8.1c) *me* *āɣā* *gāleš* *bie*
1SG.GEN mister galesh be.PST.3SG
'My husband was a galesh' (Ziyarat Mazandarani, Shokri et al. 2013:87)

Structurally equivalent clauses in which "be/become" copular verbs are accompanied by two unflagged NPs occasionally also express other domains of nominal predication. Both clauses in (8.2a-b), from Gorani and Kupia respectively, express the predicate locative functional domain with unflagged NPs encoding the ground element. More frequently, ground elements are encoded by NPs flagged by some locative marker. Similarly, examples (8.3a-b), from Middle Persian and Darai, express the predicative possession functional domain and in both, the possessor and the possessed arguments are encoded by unflagged NPs. Such situations, in which two or more domains of nominal predication are expressed by the same configurations of coding means in a single language, are referred to in this chapter as instances of "co-expression".

(8.2a) *usā* *āsā* *faransa* *biya*
master then France be.PST.3SG
'At that time, the master was in France' (Gorani, Mahmoudveysi et al. 2012, 4:65)

(8.2b) *aŋgrar* *santa* *kilagaḍa* *as-e*
Tuesday market Kilagada be-PRS.3SG
'(there's) a Tuesday market in Kilagada' (Kupia, Christmas & Christmas 1973a:18)

(8.3a) *harw kas ciš=ē ast*
every person thing=INDEF be.PRS.3SG
'Every person has one thing (which is dearer than other things)' (Middle Persian, DK6 33)

(8.3b) *ukʰ=rə dulhi=k cʰawa nidz=bʰəi-lə*
3SG=GEN wife=3SG.POSS son NEG=be-PST
'his wife did not have a son (lit. his wife was not son)' (Darai, Dhakal 2013:79)

The expression of nominal predication domains in the data is not limited to clauses with cognate "be/become" verbs. Across the languages analyzed here, other copular elements (verbal or otherwise) may express nominal predication, in addition to copula-free clauses. In several Iranian and Indo-Aryan languages motion and location verbs have developed copular functions, and are often found in clauses expressing core nominal predication domains. In some Iranian languages, predicative possession is expressed by clauses with "have" type verbs, often alongside the older construction type in which predicative possession is expressed by clauses with "be/become". Such a wealth of structural coding means that express nominal predication could potentially allow speakers to "reserve" a unique copula (or an entire clause construction) for expressing distinct nominal predication functional domains. Thus, speakers should be able, at least in theory, to minimize if not completely eliminate, any co-expression patterns.

Examples (8.1 – 8.3) demonstrate, however, that co-expression patterns are still very much evident in naturalistic usage. But these patterns are not limited to the specific copulas used in (8.1 – 8.3) or to these functional domains. One finds, for examples, that configuration of coding means usually expressing predicate locatives are co-opted to

express proper inclusion or predicate property (e.g., the metaphoric extension of the locative phrase to express a state, as in: *my cat is in heat*). Similarly, configurations of coding means which usually express predicative possession may also express the predicate locative or the existential functional domains (e.g., *the field has lots of weeds*). Such co-optations have been often given as evidence for the basic similarity between possessors and locations, or between states and locations (e.g., Jackendoff 1983, Baron and Herslund 2001, DeLancey 2000, *inter alia*).

In light of these co-expression patterns, this chapter asks two closely related questions: (a) what are the co-expression tendencies of the six nominal predication domains in Indo-Iranian? That is, to what degree does the relationships between locations and states or possessors and locations, affect the co-expression of different nominal predication domains? and (b) can we measure the degree to which nominal predication domains are co-expressed in a way that permits crosslinguistic comparison of co-expression patterns? This chapter gives an affirmative answer to the second question by proposing such a measure and using it to analyze the patterns of co-expression in the nominal predication domain, thus answering the first question as well.

In Section 8.1, I re-introduce the definitions used in this study for the six nominal predication domains, and briefly review some of the literature dealing with co-expression in the nominal predication domain. In Section 8.2, I present the data used for this chapter, which is a subset of the full set of corpora used for the entire study, and the method of token collection and coding. Then, in Section 8.3, I measure the co-expression patterns based on how close the relationship between two sets of configurations of coding means used to express two distinct domains is to proper set inclusion. After this, I use this

measure to analyze some co-expression patterns of nominal predication domains, replicating parts of two published papers (Clark 1978, Stassen 2013a). The chapter ends with a brief summary and conclusions.

8.1 Background

8.1.1 Previous Co-Expression Studies

Chapters IV through VI of this dissertation are semasiological in nature in that they describe and analyze the ways in which a set of cognate forms, or constructions, are used, and group language specific constructions into construction types. The current chapter, in contrast, is onomasiological in nature in that it compares the grammatical means used to encode a field of six functional domains. Such an approach is common in typological and comparative studies, where some functional domain is defined (e.g., as a "comparative concept," following Haspelmath 2010, Croft 2016, see Chapters I and III), and the means used to express it are compared across a sample of languages.

Several studies have taken a similar approach to comparing the co-expression patterns of different nominal predication domains. Here, I will briefly review two such studies: Stassen's 2013a World Atlas of Language Structures (WALS) entry on the co-expression of core nominal predication and predicate locative, and Clark's 1978 study of the expression of predicative possession, the predicate locative, and the existential (in her terminology: "locatives"). Some of the analyses reported in these two studies are replicated in this chapter using the Indo-Iranian sample the methods presented in this chapter.

Stassen's (2013a) WALS entry compares the expression of the predicate locative to what he calls "nominal predication" which includes equation, predicate property, and proper inclusion. I refer to these three functions as "core nominal predication". Stassen asks whether constructions expressing core nominal predication and the predicate location functional domain make use of the same copular elements or not. Thus, Stassen restricts the study of co-expression in that the only structural variable considered is the identity of the copula and the range of possible answers is binary.

Stassen notes, however, that for several languages in his sample a binary answer is not entirely satisfactory (see also Stassen 1997). These are languages in which the sets of copulas used for the predicate locative and core nominal predication partially, but not entirely, overlap. In the ten languages analyzed for this chapter, and the sixteen analyzed for this dissertation, cognate "be/become" verbs are often express both domains, and copulas grammaticalized more recently (e.g., from motion or posture verbs) often have a more restricted usage in the number of functional domains they express. Thus, in the languages analyzed here, the sets of copulas used in clauses expressing these domains does not overlap as well.

Stassen's solution for this situation is to restrict the types of copulas he analyzes to those that do not interact with semantics of time stability or permanency (see Givón 2001, Stassen 1997). This is an ad hoc solution, even if theoretically motivated, which has the advantage of fitting the possible answers to Stassen's main question into the WALS categorical coding system, which does not allow for continuous coding. Other linguistic atlases such as the *Atlas of of Pidgin and Creole language structures* (APiCS, Michaelis et al. 2013) do allow for continuous coding, at least in some entries.

Another problem with setting limits on "eligible" copulas is that it leads to a situation in which there are two distinct types of languages which are coded in Stassen's 2013a WALS entry as not co-expressing predicate locative and core nominal predication. The first type is composed of languages that really never use identical copulas to co-express the two domains. The second type is composed of languages that do use the identical copula(s) to co-express the two domains, but do so with copulas interacting with time-stability semantics. By limiting possible intralinguistic variance in his coding, Stassen 2013a ends up underestimating the amount of co-expression in the data.

Finally, the scope of "co-expression" in Stassen 2013a includes only the identity of the copula. Other structural coding means, such as the flagging of NPs or relative word order, are excluded. In the Indo-Iranian languages analyzed here, all six nominal predication domains are expressed by cognate "be/become" verbs (at least marginally), but also differ from each other with respect to coding means such as flagging of NPs and relative word order. Thus, the typology of co-expression in the nominal predication domain might seem misleadingly flat if we limited the structural variables studied to the copula alone.

In examples (8.4a-f), from Middle Persian, the copula type is identical. Examples (8.4a-c) encode core nominal predication; example (8.4e) encodes predicate locative. Under the coding criteria of Stassen 2013a, these examples would lead to the classification of Middle Persian as language in which the predicate locative and the core nominal predication domains share the same coding means. Any analysis of (8.4a-c) and (8.4e) would have to point out, however, that while the predicate in (8.4a-c) is expressed by an unflagged NP, the predicate in (8.4e) is expressed by a prepositional phrase headed

by the locative preposition *pad* ‘in, at, by’. There is a grammatical difference, then, in coding properties between (8.4a-c) and (8.4e), and these two sets of examples do not share the exact same configuration of structural coding means. Similarly, the coding properties in (8.4d, f) differ from those in (8.4a-c, e) and cannot be analyzed as exhibiting co-expression. In sum, examples (8.4a-f) use at least four distinct configuration of structural coding means to express different nominal predication functional domains.

(8.4a) *dēn ān baw-ēd*
religion DEM be.PRS-3SG
‘religion is this’ (Middle Persian, DK6 36)

(8.4b) *pašēmān baw-ēd*
regretful be.PRS-3SG
‘(after he does this evil deed) he is regretful’ (Middle Persian, DK6 71)

(8.4c) *ōy az harw dō ōh baw-ēd*
DEM from all two DEM be.PRS-3SG
‘he is one of those two’ (Middle Persian, PRDD 4)

(8.4d) *čēon=šān xwadāy ud dahibed ud sālār ud dastwar ī dēn-āgāh*
as=3PL lord and ruler and leader and priest LNK religion-knowledge
ne būd
NEG be.PST.3SG
‘because they had no lord, ruler, leader, or priest well versed in the religion’
(Middle Persian, AWN 1.8)

(8.4e) *dušmenān pad rāh būd h-ēnd*
enemy-PL by road be.PRS be.PRS-3PL
‘the enemies are on the road’ (Middle Persian, PRDD 10)

(8.4f) *būd* *dastwar kē=š* *ēdōn guft*
be.PST.3SG priest REL=3SG thus say.PST.3SG
'there was a priest who said thus:' (Middle Persian, PRDD 15.e)

In sum, examining the copula alone in its scope of co-expression over-simplifies the grammatical situation (see also Payne 2009, for other languages). To avoid such simplification, the analysis in this chapter is based on entire configurations of coding means including the type of copula used, or the lack of copula; the number and flagging of accompanying elements; indexation of these elements on verbal copulas (where applicable); and the relative word order of the elements.

Clark 1978 compares the expression of what she calls "locatives": predicative possession, the predicate locative, and the existential. She considers several structural coding means, most notably the copula and the relative word order of the topic and the predicate. She does not, however, consider these coding means as an ensemble, but compares the various pieces separately. In contrast, in this chapter I compare configurations of coding means as an ensemble, and then attempt to isolate the coding means which drive co-expression patterns.

To illustrate Clark's methodology, in one section of her study she asks whether the expression of the predicate locative and the existential functional domains differs in the relative word order of the figure and the ground. She finds the answer is affirmative and argues that for the most part, in clauses expressing the existential, that relative word order would be ground – figure, and in clauses expressing the predicate locative, figure – ground. Her finding that in many languages the difference between the predicate locative

and the existential correlates with relative word order is re-affirmed in this chapter but by using a different method and different data.

There are, however, several instances in Clark's sample and in the data analyzed for this chapter where the predicate locative and the existential are distinguished by means other than word order. This is demonstrated in examples (8.5a-b), from Modern Hebrew.[85] Both examples feature the Modern Hebrew predicative elements Clark 1978 mentions for the predicate locative and the existential: *yɛʃ* 'exists' for (affirmative) existential clauses and predicative possession, and *nimtsa*[86] 'be located' or a verbless clause for predicate locative clauses. In (8.5a-b), the ground element is expressed by the PP *baʔaʁɔn* 'in the cupboard' and the figure is expressed by the NP (*ha*)*ʔɔʁɛz* '(the) rice'. The relative order of the figure and the ground in both examples is identical: The figure precedes the ground. While in (8.5a) the figure directly precedes the ground, in (8.5b) the verb form *nimtsa,* the passive-medial form of "find", is potentially deployed between the figure and the ground. When this happens, the word order in (8.5b) is a result of the S-V word order in Hebrew intransitive clauses. Thus in (8.5a-b), the order of the figure and the ground (Clark's 1978 variable) is identical.

[85] Clark does not include examples in her paper. I have drawn up the Modern Hebrew examples here based on the Modern Hebrew copulas and word order reported by Clark.

[86] Clark lists this verb for clauses expressing the predicate locative. It should be mentioned that this is a finite verb form, the passive-medial form of "find". Thus, examples like (8.5b) are intransitive verbal clauses, and (8.5b) literally translates as "the rice is found in the cupboard" (cf. Turkish *bulunur* 'is found' from *bul-* 'find', used to express existence in a given location).

(8.5a) *yɛʃ ʔʁɛz b=a-ʔaʁɔn*
EXIST rice in-DEF-cupboard
'there is rice in the cupboard' (Modern Hebrew, constructed example)[87]

(8.5b) *ha-ʔɔʁɛz (nimts-a) b-a-ʔaʁɔn*
DEF-rice (find.MID-3MSG) in-DEF-cupboard
'the rice is in the cupboard' (Modern Hebrew, constructed example)

In Modern Hebrew, then, relative word order of the figure and the ground is not a reliable cue for distinguishing between predicate locative and the existential; the two can be distinguished from each other by the type of copular element. In existential clauses in Modern Hebrew the copula is *yɛʃ,* or its negative counterpart *ʔɛn*, whereas the predicate locative is encoded by intransitive clauses with the finite verb form *nimtsa* or by verbless clauses.

8.1.2 Definitions

Perhaps unsurprisingly, different studies use slightly different definitions and terms for the functional domains subsumed under nominal predication. The distinct approaches and definitions result in differences in the number of nominal predication domains and in variation in their scope. Here I follow the definitions found in Stassen 1997, Payne 1997:111-113, and Creissels 2013, 2014, given earlier in Chapter V. To summarize, following Clark 1978 and Stassen 1997, 2013a, among others, I split nominal

[87] Out of context this sentence is interpreted as an existential. However, one could imagine it as a possible, even if slightly odd, answer to the question "Where is the rice?". This ambiguity shows the importance of using contextualized data, through either directed elicitation or naturalistic texts.

predication in two: core nominal predication and Clark's "locatives". Core nominal predication is composed of equation, predicate property, and proper inclusion. Equation, also often called "equative-identificational", expresses the equivalence in reference between two phrases. Predicate property, often also called "predicate attribute", predicates that the referent of one phrase has a property expressed by another phrase.[88] Finally, proper inclusion predicates that the referent of an expression belongs to a set which includes other similar entities as well.

What Clark 1978 calls "locatives" is composed predicative possession, the predicate locative, and the existential. Predicative possession, defined here after Payne 2009 (who follows Langacker 1993), is expressed by clauses predicating a control relationship (e.g., ownership) or an intimate relationship (e.g., kinship) between the referents of two phrases. This definition would encompass most uses of English *have*, French *avoir,* or New Persian *daʃtan* as main verbal predicates.

The predicate locative and the existential are defined here, following Payne 1997:111-113, Creissels 2013, 2014, as two different construal options of the same situation. The predicate locative is defined as predicating the location of a figure in, spatial or temporal, relationship to a ground. The existential, on the other hand, expresses the existence of a specific figure in some, often specified, location.

The definitions used here generalize over some semantic differences that other comparative studies might wish to draw. For example, the predicate property domain

---

[88] In this study, I prefer the term "predicate property" over "predicate attribute" to avoid the association of lexical adjectives with the predicates in this domain. While in many languages, including those analyzed in this chapter, lexical adjectives are frequently used as predicates in clauses expressing predicate property, their role there is not unique. For example, in the English clauses *I am an asshole* and *my wife is a genius!,* both of which predicate properties, the predicates are lexically nouns, as evidenced by the indefinite article).

does not distinguish between permanent and temporal properties (*she is smart* vs. *she is hungry*). In addition, the definition of predicative possession does not distinguish between types of possession which are sometimes distinguished grammatically. Making these more delicate distinctions (e.g., following Hengeveld 1992) might, I believe, be profitable only if much more data from each language is analyzed.

8.2 Data, Coding Scheme, and Method

8.2.1 Data

As described in Chapter II, the data for this study comes from published naturalistic texts in sixteen Indo-Iranian languages. The main, practical, reason for this choice is that, for the most part, grammatical descriptions and sketches do not discuss the construction types of the cognate verbs compared here. The analysis of nominal predication is also often played down in grammars and grammatical sketches of Indo-Iranian languages. Many grammars and grammatical sketches ignore nominal predication altogether, while others only briefly describe it in a few pages and a handful of examples (e.g., Kachru's 2006 Hindi grammar, where only six pages are dedicated mostly, but not exclusively, to the grammar of nominal predication). To better understand the degree to which the six nominal predication domains are co-expressed, this chapter uses the same published naturalistic texts that were used in the preceding chapters.

The set of languages analyzed here is ten of the sixteen languages for which data was collected. The texts from these languages were manually scanned for instances of clauses expressing nominal predication with cognate "be/become" verbs or with other verbal or non-verbal elements functioning as copulas. This cannot be done by using an

automated search for specific verb forms, as in many languages there are instances of verbless clauses expressing at least a some of the six nominal predication domains.[89] It is also impossible to assess in advance exactly which verbs are used as copulas in any specific language, particularly motion and posture verbs. The amount of labor this task demands is one of the reasons for the smaller set of languages considered in this chapter. For each of the ten languages analyzed for this chapter, I have been able to identify 200 to 500 clauses expressing nominal predication.

8.2.2 Coding

The clauses expressing nominal predication were arranged in ten tables, one for each language. Each example was coded for a set of functional and structural variables. This coding forms the basis for the analysis in this chapter. In more than one sense, the coding of examples *is* the analysis, and the tests and measures presented later in this chapter are merely tools for aggregating and presenting this analysis in a readable format.

The only functional variable that these examples were coded for is the functional domain of nominal predication they express: "equation", "predicative property" (abbreviated as "property"), "proper inclusion" (abbreviated as "inclusion"), "predicative possession" (abbreviated as "possession"), "predicate locative" (abbreviated as "locative"), and "existential".

Often, it was impossible to decide without context which functional domain is expressed by a specific clause. The oft-cited clause *she is a teacher*, for example, can

[89] Not to mention that the functional range of many of the copulas across Indo-Iranian is broader than nominal predication alone.

express proper inclusion (she is a member of the set of professional teachers) or predicate property (she explains well and has the talent of a teacher, while in fact she's a student). For some clauses, the only way to reliably decide whether a specific clause expresses one type of nominal predication versus another involves considering it in its context. Only about 2% of the examples proved difficult to analyze even in context; these were omitted from subsequent analysis.[90]

Apart from the functional domain it expresses, each example was also coded for several structural coding means. These include the type of copula used in the clause if there was one, the flagging of the different arguments, the indexation of arguments on any copula, and the relative word order of the arguments and the copula. The possible values for the copula variable are the dictionary form of the copula (usually the infinitive form) which is often abbreviated here as the first phoneme of the copula (e.g., "*bʰ*" or "*h*"). Verbless clauses received the value "none" in this slot.

The different non-verbal components in nominal predication clauses are expressed by unflagged NPs, NPs flagged by some case marker or an adposition, or zero. Components expressed by unflagged NPs were assigned the code "NP". Those flagged by some case marker were assigned a code based on the name traditionally used for that case marker in grammatical descriptions of that specific language (e.g. "dative" or "oblique"), and those flagged by a specific adposition were assigned a code based on that adposition (e.g., *pad* for Middle Persian preposition *pad* 'in, at, by' or *par* for Hindi post-

[90] For the most part, these examples were analyzable as expressing predicate property or proper inclusion, or as expressing predicate locative and the existential. Their inclusion would not have affected the results presented below, as they would not add any information otherwise not included.

position *par* ‘on’). Finally, when a zero mention type was found, “zero” was given as a code.

Two further variables were coded for: the relative word order of the two arguments and the copula, and the indexation on the copula. In clauses where one of the arguments was expressed by a zero, it was not included in the relative word order. In clauses where the copula did not index any argument (verbless clauses, non-verbal copulas, participle forms of copular verbs), the code “none” was assigned.

To illustrate this coding schema, consider examples (8.6a-c), from Middle Persian. All three examples express predicative possession, and in all three the possessed argument is expressed by an unflagged NP indexed on the verb. This is reflected in the identical coding of these variables for each of (8.6a-c) in Table 8.1. The three examples differ in the type of main verb used and the flagging of the possessor argument. In (8.6a, c) the main verb is *būd*, and hence “B” was recorded for both examples for the verb type variable. In example (8.6b), where the main verb form is a form of the *hē* type copula, “H” was recorded. For the possessor argument, “clitic” was recorded for (8.6a), NP for (8.6b), and the type of adposition used, *rāy*, was coded for example (8.6c).

(8.6a) *čēon=šān xwadāy ud dahibed ud ... ne būd*
as=3PL lord and ruler and … NEG be.PST.3SG
‘they did not have any lord, ruler…’ (Middle Persian, AWN 1..8)

(8.6b) *harw kas čiš=ē ast*
every person thing-INDEF be.PRS.3SG
‘every person has a thing’ (Middle Persian, DK6 33)

(8.6c) *ud ōy wirāz rāy haft xwah būd h-ēnd*
and DEM wiraz to seven sister be.PST be.PRS-3PL
'and Wiraz had seven sisters' (Middle Persian, AWN 2.1)

Table 8.1: the coding of examples (8.6a-c)

Function	Verb type	ARG1	ARG2	Indexing	Word Order
Possession	B	NP	Clitic	1	Clitic-NP-v
Possession	H	NP	NP	1	NP-NP-v
Possession	B	NP	Rāy	1	Rāy-NP-v

Examples (8.6a-c) illustrate three distinct configurations of structural coding means expressing the same nominal predication domain. As such, they cannot show any co-expression pattern. In contrast, examples (8.7a-b), from Gorani, repeated from earlier in this chapter, do show a co-expression pattern. In both examples, the same verbal copula is used, accompanied by two unflagged NPs. Example (8.7a), however, expresses equation whereas (8.7b) expresses the predicate locative.

(8.7a) *īrij kuř=aš biya*
irij son=3SG be.PST.3SG
'Irij was his son' (Gorani, Mahmoudveysi et al. 2012, 5:8)

(8.7b) *usā āsā faransa biya*
master then France be.PST.3SG
'At that time, the master was in France' (Gorani, Mahmoudveysi et al. 2012, 4:65)

In this chapter, I refer to the set of structural coding means used in a clause as the "configuration of structural coding means". Thus, instances of co-expression of different

nominal predication domains are instances where two clauses have the same configuration of structural coding means but express distinct domains.

8.2.3 Visualizing and Measuring Similarity in Co-Expression Patterns

The result of applying this coding process to data from the ten languages of the sample is ten tables with the same six columns as in Table 8.1, and as many rows as the number of clauses expressing nominal predication found. Using an R script, I converted these tables to matrices with six rows representing the six nominal predication domains, and as many columns as there are *distinct* configurations of structural coding means. Each cell in these matrices represents whether a particular configuration of structural coding means expresses a particular nominal predication domain. If the configuration of structural coding means represented by that column is used to express the domain represented by the row, then the cell has the value of 1; otherwise, it has the value of 0. Instances of co-expression are instances where in the same column there is more than one cell with 1, or alternatively, when the sum of a specific column is bigger than 1. If a specific configuration of structural coding means expressed only one type of nominal predication, then there would be a single cell with the value of 1 in that column.

The matrices produced in this way can be visualized as bipartite network graphs (Wasserman & Faust 1994:299-302; also sometimes called bimodal graphs of bimodal networks). In regular network graphs (also called uni-modal or uni-partite networks), each node can potentially be connected to all other nodes by edges. Thus, all nodes in such networks represent items of the same type, or at least of comparable types. In social network analysis, each node in a uni-modal network can be used to represent a person, or

a member of the community, and two nodes are connected by an edge if the two people they represent interact socially.

Bipartite networks, on the other hand, are networks in which nodes can be divided in two sets, such that members of one set can only connected to members of the other set. Thus, nodes in bipartite networks can be used to represent two qualitatively distinct types of entities which interact in some ways. In social network analysis, bipartite networks are used to study the interactions among members of a community in social institutions or gathering. That is, one type of node can be used to represent members of the community, and the second type cam be used to represent classes they take, parties they might go to, and so on.

Here, I use bipartite networks to visualize the relationship between the six nominal predication domains and the configurations of structural coding means used to express them. In the bipartite networks presented here, the first type of nodes represents the nominal predication functional domains while the second represents the different configurations of structural coding means. If a configuration of structural coding means expresses some functional domain, the two nodes representing them are connected by an edge; otherwise, the two are not connected. Nodes representing configurations of structural coding means that co-express several domains, then, are connected to more several nodes that represent distinct nominal predication domain.

To construct this visualization, I wrote a short R script using the iGraph R package (Csardi & Nepusz 2006), which constructs bipartite networks from tables like the one described above. In plotting the network, the iGraph package implements the Fruchtman-Reingold algorithm (Fruchterman & Reingold 1991), which tries to make all

nodes and edges visible and non-overlapping on the one hand, while also minimizing the overall length of edges while keeping them observable. This means that as the lengths of edges are minimized, nodes representing domains which are more commonly co-expressed will be plotted closer to each other. In the visualization of the bipartite networks presented here, the nodes representing the six nominal predication domains are large, and are labeled by names of the domains they represent. The nodes representing the different configurations of structural coding means are small, and are colored based on the copular element in them (nodes with the same color use the same copula).

Figures 8.1 and 8.2 illustrate two extreme types of such networks. Figure 8.1 represents an extreme situation in which all six nominal predication domains are co-expressed by the same set of configurations of structural coding means. Hence, all nodes representing the nominal predication domains are connected to all nodes representing configurations of structural coding means.

Figure 8.2 represents the exact opposite situation in which none of the nominal predication domains shares any configuration of structural coding means. The configurations used to encode each domain differ in at least in one variable from those expressing other domains. In the bipartite network, none of the nodes representing structural coding means is connected to more than one node representing a nominal predication domain.

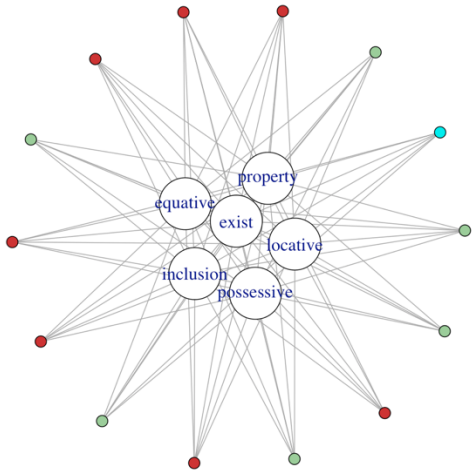


Figure 8.1: All functional domains are co-expressed

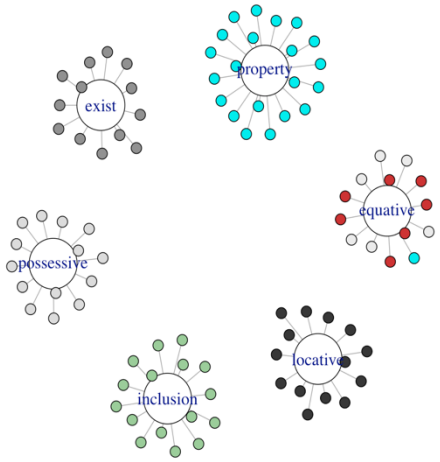


Figure 8.2: No functional domains are co-expressed

The situations illustrated by Figures 8.1 and 8.2 are extreme situations which are not found in any of the ten languages analyzed here, and probably are not found in any natural language. Figures 8.3 and 8.4 illustrate the situation as recorded for Gorani and Kotia Oriya. In both languages, the nodes representing core nominal predication domains are closest to each other, suggesting that the sets of configurations of structural coding means expressing these domains are more similar to each other than to the sets of

configurations used to express other domains. Figures 8.3 and 8.4 also illustrate some differences in the co-expression patterns in Gorani and Kotia Oriya. Consider the relative distance between the nodes representing predicative possession, the predicate locative, and the existential. In Gorani, the set of configurations of coding means expressing both is small, and the two are plotted on opposite sides of the bipartite network. Predicative possession, however, is often co-expressed with the existential, and less often with the predicate locative or other domains, and is thus plotted between the two.

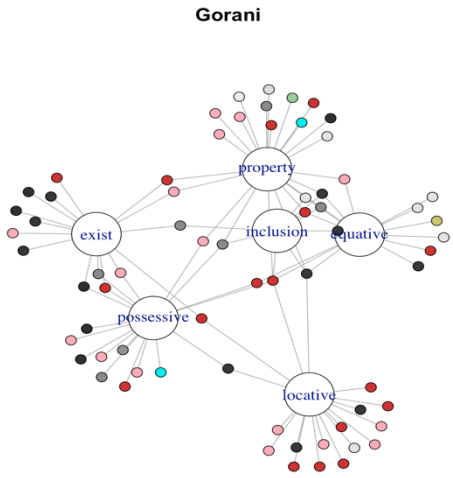


Figure 8.3 Co-expression of nominal predication in Gorani

In contrast, the co-expression patterns of these three domains in Kotia Oriya, visualized in Figure 8.4, is completely different. Here, the predicate locative and the existential are co-expressed by several configurations of structural coding means, and are either seldom or never co-expressed with predicative possession. Thus, in Figure 8.4, the predicate locative and the existential are plotted relatively close to each other, while predicative possession is plotted further away. The co-expression patterns of predicative possession, the predicate locative, and the existential found in Kotia Oriya and Gorani are

of two different types. In Gorani, predicative possession is similar to both other domains, while in Kotia Oriya, predicative possession is completely distinct from them.

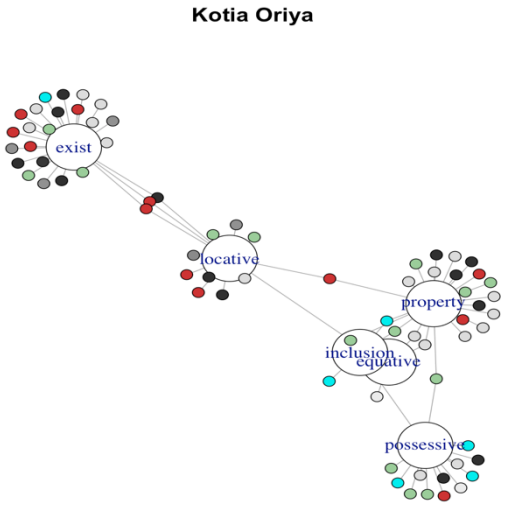


Figure 8.4 Co-expression of nominal predication in Kotia Oriya

Bipartite networks, illustrated by Figures 8.1 – 8.4, provide a nice visualization of the data and assist in formulating hypotheses about the similarity between different nominal predication domains in terms of the coding means used to co-express them. It is difficult, however, to rely solely on these visualizations for a direct answer to questions about the degree of co-expression across languages and at the same time, there are many different ways to define, and consequently to measure, how similar are two sets of configurations of structural coding means.

Many linguists have used Euclidean or Hamming distance to measure the similarity between two or more languages relative to some variable. Euclidean and Hamming distances are used in Chapter IX for comparing the grammatical behavior of specific lexical items in different languages, and were also used in Chapter VII for comparing constructions of "do/make" in terms of the event types they express. The similarity measured in this chapter, however, is not between two languages, but between

pairs of nominal predication domains in terms of the sets of configurations of coding means used to express them. I will briefly argue here that Euclidean or Hamming distance are not good methods to measure how similar are the coding means expressing pairs of distinct functional domains.

In Euclidean and Hamming distance measures, shared absences are treated as similarities. In the context of this chapter, this means that if neither of two domains is expressed by some configuration of coding means, then under Euclidean or Hamming distance this would contribute to their overall similarity. The contribution of shared absences to overall similarity in the type of data used here can, and will, quickly outweigh the contribution of shared occurrences (i.e., actual co-expression of two domains). Thus, for example, if in a given language, there are many unique ways to express predicate property, this will increase the number of shared absences between the existential and predicate locative, which in turn would decrease the Euclidean or Hamming distance between the two.

Consider the following constructed example. In language A, predicative possession and the existential do not share any configurations of structural coding means. In language B the two share three configurations of structural coding means. Thus, under the usual understanding of similarity (only counting actual instances of co-expression), the existential and predicative possession are grammatically completely distinct in language A, but are at least somewhat grammatically similar in language B.

There is, however, a small twist. In both languages, predicate property is expressed by several configurations of structural coding means: in language A by fifteen configurations, and in language B by eight. None of these configurations is also used to

express the existential or predicative possession. In this new situation, using Euclidean or Hamming distance, the existential and predicative possession would be more similar in language A than in language B, because in language A there are fifteen instances of shared zeroes, i.e. shared absences, and in language B only eight. This is despite the fact that in language A the existential and predicative possession are never co-expressed by any configuration coding means. This goes against the usual understating of grammatical similarity of functional domains, which should be based only on the number of attested shared configurations of coding means, i.e., only on shared occurrences.

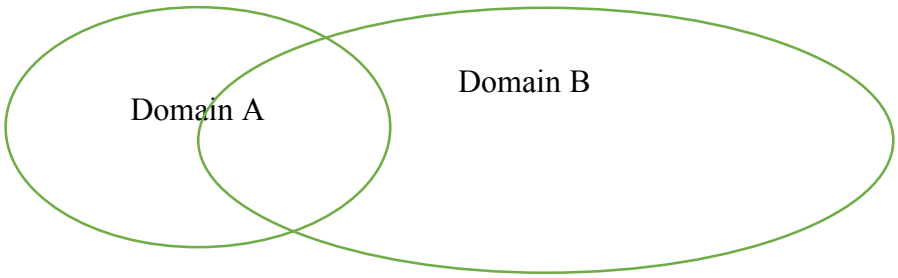
Instead of using Euclidean or Hamming distance, I propose here to measure the similarity between two functional domains by calculating how close to complete set inclusion the relationship is between the sets of configurations of coding means that express both domains. The closer this relationship is to complete set inclusion, the more grammatically similar the domains are. The following equation operationalizes this measure of similarity between two domains of nominal predications, A and B: the number of shared configurations, i.e., configurations used to express both domains A and B is divided by the size of the smaller set of configurations of the two sets expressing each domain. For reasons of convenience, I treat this as a measure of dissimilarity, so I subtract the result of this division from 1.

The result of this calculation is a number between zero and one. If the result equals zero, this means that the smaller set of configurations is completely included in the larger one. If the result equals one, this means that the two sets of configurations are completely disjoint. The closer to zero this result is, the closer the relationship between the two sets expressing domains A and B to complete set inclusion.

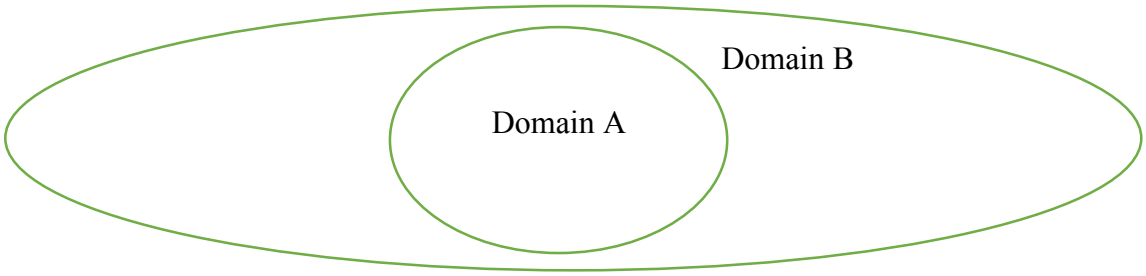
$$Dissimilariry\ (A,B) = 1 - \frac{\#\ \text{of shared configurations}}{MIN\ (\#\ \ expressing\ A, \#\ \ expressing\ B)}$$

Consider, for illustration, the three scenarios depicted by the Venn diagrams below. The ovals represent the sets of configurations of coding means that express two domains, A and B. In scenario 1, the two domains are not co-expressed by any configurations, and the measure of dissimilarity proposed above is equal to 1. In scenario 2, the two domains share some configurations of coding means, so the result of applying the calculation above would be a number somewhere between one and zero. In scenario 3, the set of configurations used to express domain A is completely included in the set of configurations used to express domain B, so the result for applying the above formula to this situation would be 0.

Scenario 1: disjoint sets



Scenario 2: partially joint sets



Scenario 3: complete set inclusion

The two tools presented in this section, the visualization of the expression of nominal predication as bipartite networks and the measure proposed for dissimilarity, are used in the next section below to study co-express patterns of the six different nominal predication domains.

8.3 Co-Expression of Nominal Predication Domains

8.3.1 Predicate Locative and Core Nominal Predication

This section asks a similar question to that posed in Stassen's 2013a WALS entry: are predicate locatives encoded by the same configurations of coding means as core nominal predication? The question is answered by using the formula presented in Subsection 8.2, and by taking into account entire configurations of structural coding

means. Another difference between this section and Stassen's 2013a WALS entry is the size of the sample and the type of data. Stassen's study was based on a genealogically balanced sample of 386 languages and data was collected from published descriptions, mostly in grammars. The study here is based on published textual data from ten Indo-Iranian languages: the sample is much smaller, and the data comes from a specific genealogical family.

Examples (8.7a-b), repeated here as (8.8a-b), have already shown that even when we consider entire configurations of structural coding means instead of just subsets of them, we can still identify clear instances of co-expression of the predicate locative and core nominal predication. Examples (8.8a-b) represent one type of situation which leads to co-expression of the predicative locative and core nominal predication, in which the locative ground is expressed by a semantically locative noun in an unflagged NP.

(8.8a) *īrij kuř=aš biya*
irij son=3SG be.PST.3SG
'Irij was his son' (Gorani, Mahmoudveysi et al. 2012, 5:8)

(8.8b) *usā āsā faransa biya*
master then France be.PST.3SG
'At that time, the master was in France' (Gorani, Mahmoudveysi et al. 2012, 4:65)

Another type of clauses involved in the co-expression of the predicate locative and core nominal predication is demonstrated in (8.9a-b). In both examples, "be/become" is accompanied by an unflagged NP and a PP headed by *andar* 'in'. In (8.9a), the PP

expresses a location (a river), and the clause expresses predicate locative. In (8.9b), the PP expresses a state, "in doubt", and the entire clause predicates the state of the people in a particular situation, and thus expresses predicate property. In all instances of this type of co-expression, clause expressing core nominal predication expressed the predicate property or proper inclusion functional domains, and never expressed equation.

(8.9a) *was ruwān ud frawahr-ān andar ān rōd būd h-ēnd*
many soul and fravashi-PL in DEM river be.PST be.PRS-3PL
'and many souls and fravashi were in that river' (Middle Persian, AWN 16)

(8.9b) *mardōm-ān andar gumān būd h-ēnd*
man-PL in doubt be.PST be.PRS-3PL
'men were full of doubt' (Middle Persian, AWN 1.11)

In all ten languages analyzed for this chapter, cognate "be/become" verbs express the predicate locative and core nominal predication. This means that under Stassen's 2013a typology, all ten languages would likely belong to a single category of "shared" languages, i.e., languages in which the two domains are co-expressed. The method for computing dissimilarity used in this chapter shows, however, that the degree to which the predicate locative and core nominal predication are co-expressed is far from uniform even across closely related languages. Figure 8.5 shows that this degree ranges from complete dissimilarity with a value of almost 1 (i.e., the two domains are expressed by completely disjoint sets of configurations of coding means) to a value of about 0.5 (i.e., the two domains are co-expressed by the same configurations, but each domain is also expressed by configurations not used to express the other).

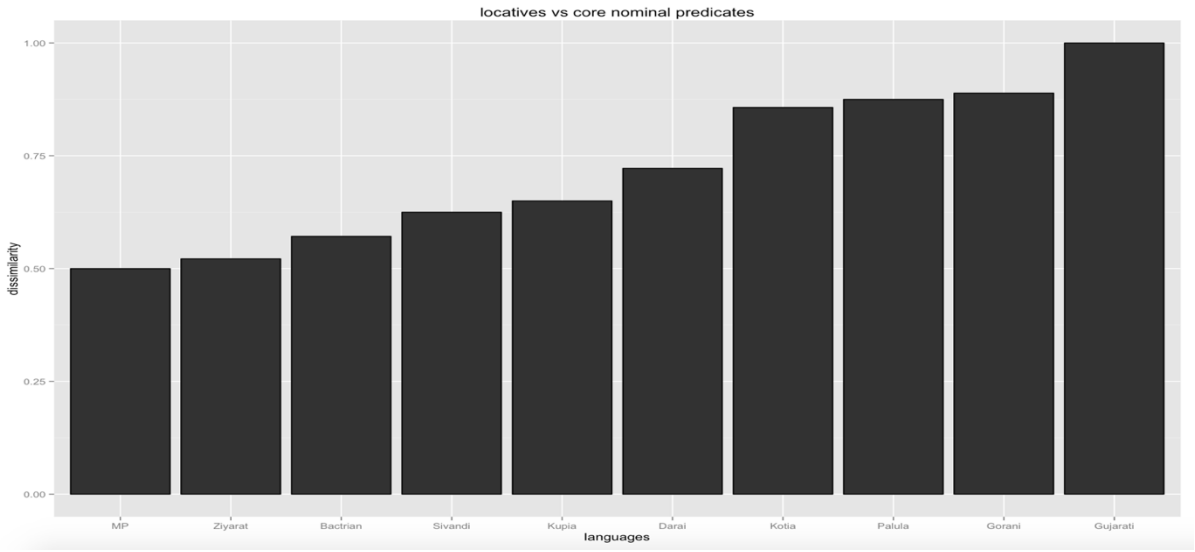


Figure 8.5: co-expression similarity of predicate locative and core nominal predication

Figure 8.5 shows, then, that Indo-Iranian languages are not uniform in the degree to which the predicate locative and core nominal predication are co-expressed by the same configurations of structural coding means, and are better described as being on a cline. On one end of this cline are languages like Colloquial Gujarati, where the predicate locative and core nominal predication are almost never co-expressed by the same configurations of coding means. On the other end are languages like Middle Persian, in which the two domains are co-expressed by a relatively large set of configurations of coding means. Thus, Indo-Iranian languages differ in the degree to which the privileged relationship between states and locations interacts with the co-expression of core nominal predication and the predicate locative functional domains.

In all languages for which some degree of co-expression is recorded, cognate "be/become" verbs participate in co-expression patterns.[91] Other copular verbs, including

[91] This includes cognates of PIE **b^{h}euH$_2$-* or **b^{h}eH$_2$u*, and **H$_1$es-* and the Indo-Aryan copulas in *(a)ch-* or *as-*.

innovative locative copulas (e.g., *daren* 'be located' in Central Iranian languages or *rehna* 'be located' in Hindi) and copulas with a motion verbs as their source, are more rarely (or almost never) participate in these co-expression patterns. Locative copulas, attested both in Iranian and Indo-Aryan are often used only in clauses expressing predicate locative, usually with the locative predicate flagged by a locative marker. Thus, the relationship between these copulas and predicate locative is tight. On the other hand, cognate "be/become" verbs are used in clauses expressing all nominal predication domains (often with different configurations of coding means, especially flagging). These verbs are then less semantically specialized, and the link between them to specific nominal predication domains is weaker than that of recently grammaticalized locative copulas. This weaker link facilitates the possible effect of the privileged relationship between states and locations, which in turn leads to a higher likelihood of co-expression by configurations in which these older "be/become" verbs function as copulas.

8.3.2 Predicative Possession and the Predicate Locative / Existential

The connection between possessors and locations, which is usually dubbed "possessors are locations", has been argued by many linguists (e.g., Jackendoff 1983, Baron & Herslund 2001, DeLancey 2000). Others have argued that the semantics of possession cannot be easily reduced to locative semantics, and that therefore possession and location are distinct semantic domains (e.g., Langacker 1993, Payne 2009). The evidence usually used to support or oppose the relationship between possession and location has to do with co-expression of the two domains by the same coding means. For example, Baron and Herslund (2001) argue that in some Germanic languages, the two domains are co-

expressed by the same means. Following Benveniste (1966), who considers French *avoir* 'have' a different construal of *être à* 'be at',[92] they argue that possession verbs in Germanic languages essentially predicate locative relations. Payne (2009), on the other hand, shows that in Maa (East Nilotic), predicative possession and the predicate locative are not co-expressed by the same coding means, especially not the same copular verbs.

Some authors involved in this debate, including Heine (1997), DeLancey (2002), and Payne (2009), have noted that there are many distinct ways in which the domains of possession and location can interact in the synchronic grammar of a specific language. In some languages, possessors are flagged by synchronically active locative markers (e.g., Tibetan; DeLancey 2000), and in other languages, constructions that commonly express predicative possession can be extended to express the predicate locative without a change in the structural coding means (e.g., *the garden has weeds*). Thus, the interaction between the grammar of a given language and the relationship between possession and location can have different effects on the expression of nominal predication functional domains in a single language as well as crosslinguistically.

In this subsection, I acknowledge that there is evidence for a privileged (i.e., cognitive or notional) relationship between possession and location, and ask how does this relationship interact with the expression of the predicate locative, the existential, and predicative possession. That is, does this privileged relationship, of whatever nature, lead to co-expression of predicative possession with either of the locative domains of the predicate locative and the existential?

---

[92] Note that following the definitions used in this study, following Creissels 2013, 2014, the predicate locative and the existential functional domains are considered to be different construals of the same situation.

It was already shown that the relationship between predicative possession and the predicate locative or the existential functional domains varies across Indo-Iranian. Consider the degree of co-expression of these three domains as visualized in the bipartite networks of Figures 8.3 and 8.4 above. In Gorani, the predicative possession node is connected to both the predicate locative and the existential nodes, and these domains are co-expressed by at least some configurations of coding means. In Kotia Oriya, in contrast, the predicative possession node is almost entirely unconnected to the predicate locative and the existential nodes; hence, these pairs of domains are never, or seldom, co-expressed by the same means in Kotia Oriya. The degree of co-expression of predicative possession and the predicate locative, and of predicative possession and the existential was calculated using the method described above and is shown in Figure 8.6.

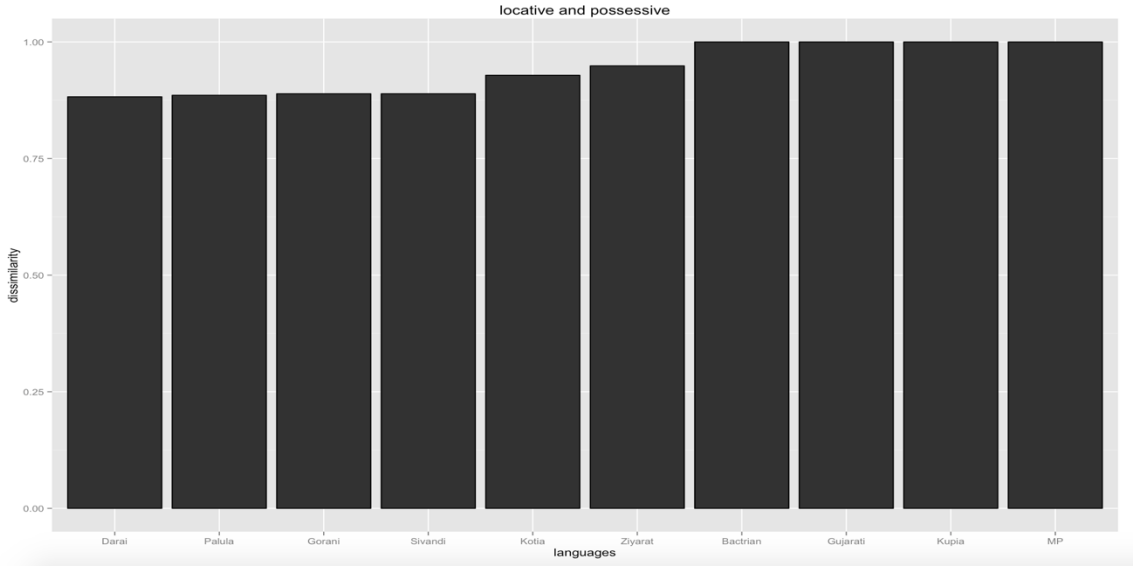


Figure 8.6: co-expression of predicative possession and the predicate locative

The dissimilarity between the expression of the predicate locative and predicative possession, shown in Figure 8.6, is rather large. In several languages (e.g., Middle Persian, Kupia, Gujarati) the two domains are never or almost never co-expressed by the

same configurations of structural coding means. In the languages in which the two domains are co-expressed, Darai, Palula, Sivandi, and Gorani, the co-expression is infrequent at best. This result suggests that synchronically, there is little evidence for an effect of the privileged relationship of possessors and location on the grammatical expression of predicative possession and the predicate locative in Indo-Iranian. The two domains are only infrequently co-expressed by the same configurations of coding means. Even in languages where the two domains are co-expressed, there are many configurations of coding means expressing only one of these domain and not the other, and this results in a low degree of overlap between the sets of configurations of coding means.

The attested co-expression patterns of the predicate locative and predicative possession are the result of two types of extensions. The first type is direct extension and a possessor flagged by some synchronically locative marker, also used to flag the locative ground in clauses expressing predicate locative. This type is illustrated by (8.10), from Palula, where the possessor is flagged by *wee* 'in'.

(8.10) *kaʈéeri bi aɽíi=wee hín-i*
knife TOP 3sg.OBL=in be.PRS-F
'he had a knife' (Palula, Liljegren and Haider 2015:117)

The second type of extension is an indirect extension that occurs when one configuration of coding means, which usually expresses neither the predicate locative or predicative possession, is co-opted to express both. Examples (8.1 – 8.3), show that the construction type in which a copula is accompanied by two unflagged NPs is sometimes

co-opted to express the predicate locative and predicative possession. When both extensions occur in the same language, the result is a co-expression pattern. As this type of co-expression requires two extensions to co-occur in the same language, it is a not very common.

In contrast to the low degree of co-expression of predicate locative and predicative possession, the interaction between the sets of configurations used to express the predicative possession and the existential functional domains is more varied. Figure 8.7 summarizes the results of applying the same measure of dissimilarity used above to the configurations that express the predicative possession and the existential functional domains. It is evident from the figure that these two domains are commonly co-expressed in some languages, such as Gujarati and Sivandi, and are never co-expressed in some languages, including Middle Persian, Kupia, and Kotia Oriya.

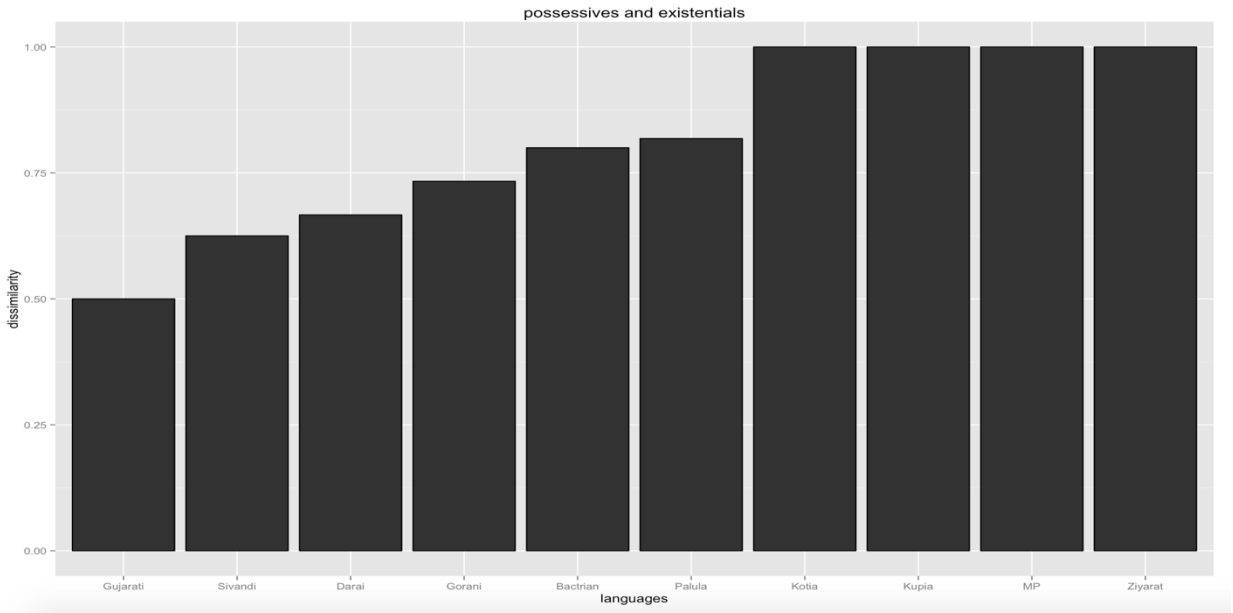


Figure 8.7: co-expression of predicative possession and the existential

The co-expression patterns in languages with lower dissimilarity rates are driven by the constructions in which possessors are expressed as adnominal modifiers of the

possessed NP. These are the constructions which follow the Heine's 1997 Genitive schema (X's Y exists). Thus, in such clauses there is only a single NP constituent accompanying the copular verb (usually a cognate of "be/become"), and this configuration is identical to that found in many existential clauses.

In those languages where predicative possession is regularly expressed by Heine's Genitive Schema, dissimilarity is driven by two factors. The first is the co-existence of other constructions expressing possession, which cannot be interpreted as an instance of Heine's Genitive Schema. This is found, for example, in Darai, where predicative possession is also expressed by constructions which belong to Heine's Locative or Companion Schema. The second cause of dissimilarity in those languages has to do with the processes of "transitivization" (also called "have drift"), and "adjectivization" (Stassen 2005, 2013a), which are illustrated in chapter V for Sivandi, Gorani, and Darai. In the data analyzed here, these processes often lead constructions which belong to Heine's Genitive schema away from it, thus give rise to configurations expressing predicative possession but not existence.

This subsection has tested the degree to which predicative possession is co-expressed with the predicate locative and the existential domains. It showed that (a) in Indo-Iranian, the privileged relationship between possession and location does not lead to high co-expression rates of predicate locative and predicative possession; and (b) that Indo-Iranian languages vary in the degree to which predicative possession and the existential are co-expressed. In some languages these two domains are seldom or never co-expressed while in others they are co-expressed frequently.

It seems that the privileged relationship between possession and location synchronically interacts to some degree with the co-expression patterns of predicative possession. For the most part, this interaction is manifested by the co-expression of the existential and predicative possession. This finding is similar to what Payne (2009) finds for Maa (East Nilotic): predicate locative and predicative possession are seldom (if ever) co-expressed, but predicative possession and the existential are, more often than not, expressed by the same verbal element.

8.3.3 Predicate Locative and the Existential

This subsection investigates the co-expression patterns of the existential and predicate locative domains. According to the definitions for predicate locative and the existential used in this dissertation, the two domains differ in the way they construe the same situation. Predicate locative clauses predicate the location of a particular element in time or space, while the existential, which, as mentioned earlier, can be thought of as "inverse-locative", predicates the existence of a particular element in some location. This subsection tests how robust the difference in coding is between the two construal options.

Figure 8.8 illustrates the co-expression patterns of the existential and the predicate locative found in the data. It shows that across Indo-Iranian there is some variation in the degree to which the two construal options are expressed by unique configurations of structural coding means: in most languages, the predicate locative and the existential are expressed by distinct configurations of structural coding means, and less commonly, the two domains are co-expressed by the same configurations of structural coding means.

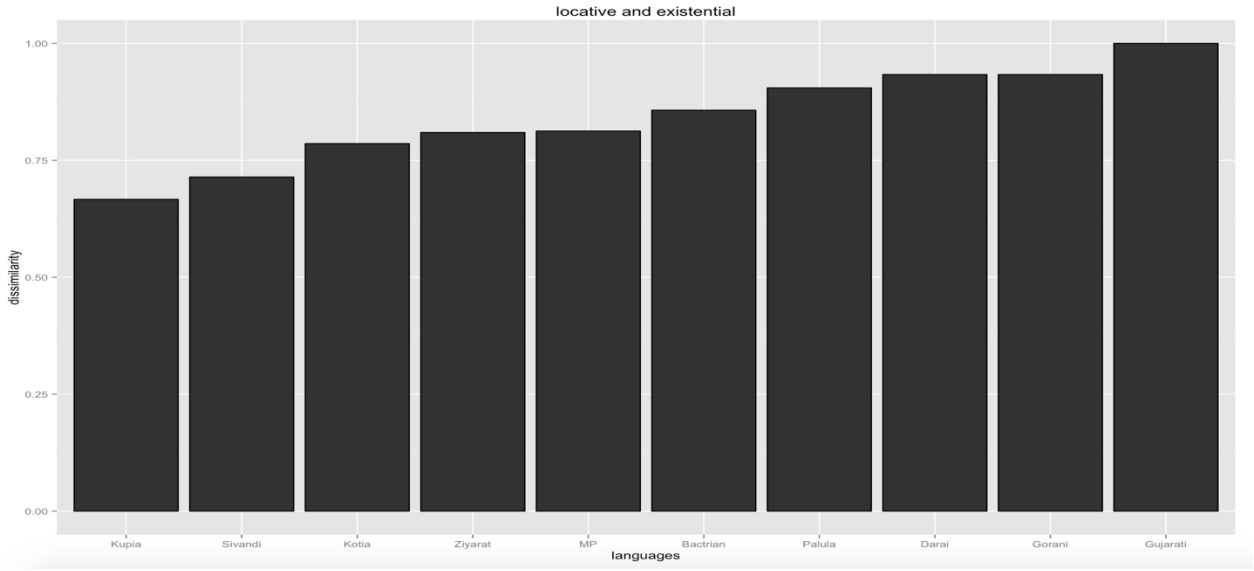


Figure 8.8: co-expression of the predicate locative and the existential

Two structural loci distinguish the existential and the predicate locative. The first is configurations of coding means in which the copula (usually the copular verbs) is accompanied by only a single NP. This construction type, which is found in all ten languages analyzed in this chapter and in all sixteen languages in the corpus, is always used to express existential semantics, and is never used to express the predicate locative. The mirror image of this situation is clauses in which the figure element is not overtly expressed, so that the only two elements overtly present are the copular verb and the locative ground, which is usually expressed by a NP flagged by a locative marker. Such clauses are rare in the data, but those that are found invariably express the predicate locative.

The second structural locus distinguishing the existential from the predicate locative is the relative word order of the figure and ground (and the copula, if present). For the most part, clauses expressing the predicate locative have a relative word order of figure-ground-verb, while clauses expressing the existential have a relative word order of

ground-figure-verb. This difference can, by definition, only arise in instances where both the figure and the ground elements are overtly expressed in the clause. This subsection, then, supports Clark's 1978 finding that the relative word order of the figure and the ground is used in many languages (in her sample of thirty languages) to differentiate the existential and the predicate locative.

8.4 Summary

Studies comparing languages using primary textual data rather than grammatical descriptions have been gaining popularity lately, but are not the norm. In this chapter, I proposed two methods for the study of co-expression in the nominal predication domain based on naturalistic data, and implemented those methods on a dataset from Indo-Iranian languages. Across this language family, "be/become" cognate verbs express the entire range of nominal predication in several distinct construction types. These construction types differ most often in the number and flagging of NPs which accompany the "be/become" verbs, and less often in the relative word order of these NPs vis-à-vis the copular verb.

It could be argued that treating configurations of coding means as ensembles will show that nominal predication functional domains are expressed by unique and dedicated grammar, and that co-expression patterns are non-existent. This chapter, however, has shown that this hypothesis is false, at least for Indo-Iranian. Even when considering entire configurations of coding means there are clear instances of different nominal predication domains co-expressed by the same grammar. To prove this point, this chapter implemented one visualization technique and one quantitative measure of dissimilarity.

First, the relationship between the six nominal predication domains and the configurations of coding means expressing them was visualized as a set of bipartite networks. These networks were used to generate hypotheses about variation and stability in co-expression patterns across Indo-Iranian. Then, a method was used which estimates the similarity in the grammar that expresses pairs of functional domains, by calculating how close the relationship of the two sets of configurations of coding means is to proper set inclusion.

The first pair of domains compared here consisted of the predicate locative and core nominal predication. It was shown that Indo-Iranian languages vary in the degree to which these two domains are expressed by the same structural coding means. Then, the same method was used to probe whether the privileged relationship between possession and location interacts with the synchronic expression of predicative possession and the predicate locative / the existential. It was found that predicative possession and the predicate locative are seldom co-expressed in Indo-Iranian, whereas predicative possession and the existential are more often co-expressed. This finding supports Payne's (2009) conclusions for Maa, where predicative possession and the predicate locative were never co-expressed, but that the existential and predicative possession do share some structural coding means.

Finally, this chapter has compared the sets of configurations of coding means that express the existential and the predicate locative. These two domains are mirror images of each other, providing distinct construal options for the same situation. Across Indo-Iranian, languages vary in the degree to which each construal option has its own unique grammar.

CHAPTER IX

HISTORICAL PATTERNS IN THE DIVERSIFICATION OF COGNATE VERB USAGE

Across Indo-Iranian, cognate "do/make", "give", and "be/become" verbs are used in many distinct construction types examined in Chapters IV through VI. These construction types are distinguished based on the function of the cognate verb (main verbal predicate, auxiliary verb etc.), the argument frame and semantics (if the cognate verb functions as the main verbal predicate), and the type of auxiliary construction (if the cognate verb functions as an auxiliary verb). These construction types vary in their crosslinguistic distribution: some are found in all, or almost all, of the languages analyzed here, while many are found only in a single language.

This chapter investigates whether there are genealogical patterns in the overall, aggregate crosslinguistic distribution of construction types in the sample of sixteen Indo-Iranian languages analyzed here. To do so, this chapter asks three nested questions. The first question is whether the aggregate usage of cognate verbs in closely related languages is more similar than in more distantly related languages. That is, is the use of the cognate verbs within subfamilies of Indo-Iranian more similar than across these subfamilies? As the answer to this question is affirmative, the second question is what is the scope of the subfamilies within which these similarities can be identified. That is, are the differences found in the level of major subfamilies (i.e., Indo-Aryan vs. Iranian) or in smaller, more recently split ones (e.g., Eastern Indo-Aryan)? Finally, since the answer to this second question varies across the three sets of cognate verbs, this chapter asks what are the diachronic processes which drive this difference.

To answer the first two questions, this chapter presents a Neighbor-Net analysis (Bryant & Moulton 2004) of the crosslinguistic distribution of construction types. Neighbor-Net is a distance based method for detecting degrees of similarity between different items (in this chapter, languages). Neighbor-Net analyses have been gaining popularity as tools in comparative and typological linguistics (e.g., Verkerk 2014, Bakker et al. 2011), and can be profitably co-opted to answer questions such as those asked in this chapter. The Neighbor-Net analysis presented in this chapter was done in both the R package Phangorn (Schliep 2011) and in SplitsTree4 (Hudson & Bryant 2006). Both produced roughly identical results. The figures presented in this chapter are taken from the SplitsTree4 output.

Section 9.1 presents the questions asked in this chapter, and situates them in the context of similar questions previously asked by linguists. Section 9.2 briefly presents the methods and the coding procedures, before presenting the Neighbor-Net analyses and their interpretation. Section 9.3, compares the different processes of change found with "do/make", "be/become", and "give", and argues that the type of processes attested with each verb interacts with the overall, aggregate crosslinguistic diversification across the family.

9.1 Introduction and Background

In the process of grammatical change, Indo-Iranian "do/make", "be/become", and "give" ended up being used in a number of distinct construction types. A glance at the crosslinguistic distribution of these patterns shows that they range from construction types found in all languages of the sample to construction types identified in one or two

languages only. Moreover, some of these construction types are attested only in languages which belong to a specific subbranch (e.g., "do/make" as an auxiliary expressing thematic continuity, found only in the Odian subbranch of Eastern Indo-Aryan), while others are distributed across branches (e.g., predicative possession in Heine 1997's goal schema).

These patterns of crosslinguistic distribution of construction types suggest that their correlation with well-established genealogical subfamilies of Indo-Iranian is far from perfect. If one wished to argue on the basis of preconceived notions that diversification of the usage of cognate verbs is completely random, one could support this notion by cherry-picking appropriate construction types based on their distribution across subbranches of Indo-Iranian. Similarly, one could cherry-pick those construction types which developed only in established subfamilies and argue for a good correlation between the usage patterns of cognate verbs and genealogical classification.[93] The analyses presented in this chapter avoid such cherry-picking and consider all identified construction types equally and without fitting them with a-priory different weights.

Questions such as the ones asked here, and especially whether the use of cognate verbs is more similar within subfamilies then across subfamilies, are infrequent in the comparative linguistics literature. There have been, however, several studies in which the main questions circle around those asked here. This subsection briefly reviews those studies and highlights the differences between the questions asked in them and in this chapter.

---

[93] Another way one could influence the result would be to assign different, a-priori, weights to different construction types, and thus increasing or decreasing the effect of different construction types on overall similarity.

In a series of studies, Verkerk (2014) investigates the encoding of motion events in data from twenty modern Indo-European languages. The data for her project comes from motion events she identified in three parallel texts (*Alice in Wonderland, Through the Looking Glass,* and *The Alchemist*) which were previously translated into the twenty languages in her sample. After collecting parallel clauses expressing motion events in her corpora, she analyzed them using well established types of motion schemata (e.g., verb-framed, satellite-framed). Using several distance-based and phylogenetic quantitative methods, Verkerk argues that there is a "historical signal" or "phylogenetic signal" in the typology of encoding motion events across Indo-European. Languages which belong to the same Indo-European branch tend to encode similar motion events by constructions of the same motion schema more often than languages which do not belong to the same branch.

The main difference between the type of questions asked by Verkerk and the questions asked here is that her questions are onomasiological in nature, asking whether the constructions expressing some functional domain do so using typologically similar constructions. In contrast, the questions asked in this chapter are semasiological in nature, comparing the different functions expressed by constructions which use some cognate lexical material. In a sense, Verkerk's questions and the questions asked in this are complementary.

Other studies have been centered around the usage of the cognate or crosslinguistically comparable lexemes in different constructions. Enfield (2003) analyzed changes in the semantics and argument frames associated with verbs which originally meant "acquire" in languages of Mainland South-East Asia, showing that these

verbs underwent a series of parallel changes, with many of them ultimately expressing ability, i.e., roughly *she acquired a tangible object → she acquired (the knowledge of) doing something → she can do something*. Enfield used the set of parallel changes attested with "acquire" to illustrate the effects of extreme language contact situations and what he calls the "epidemiology of (mental) representations". One of the goals of Enfield's study was to show that contact may result in identical pathways of change across genealogically unrelated languages. As the languages in Enfield's study are mostly not genealogically related, the different "acquire" verbs in his analysis are mostly not cognate, and the changes documented are parallel innovations, not shared innovations.

Finally, the idea that languages change at different rates is hardly new. Many Indo-Europeanists, for example, have referred to the modern Baltic languages, especially Lithuanian, as "conservative" languages, and Modern Icelandic is often treated as more conservative than its Germanic sister and niece languages. Of course, labeling languages as "conservative" in comparison to closely related languages means not that the evolution of innovative grammatical constructions is necessarily slower in these languages, but that they tend to retain more old constructions.

Several studies, however, have argued for scenarios in which the rates of parallel grammatical change differ across closely related languages. This has been shown several times for the (Italo-)Western Romance language family. In this family, French seems to be "further along" different grammaticalization pathways when compared to Italian and Spanish (e.g., Lamiroy 1999, 2003; Carlier et al. 2012 and studies therein).[94] Most of

[94] Most of these studies limit their scope to French, Italian, and Spanish, and seem to ignore dialectal variation within these languages and other Western Romance languages, such as Catalan and Portuguese.

these studies show that comparable innovative constructions in these three languages differ in their relative degrees of syntactic freedom, with French often being the most restrictive one, implying that these constructions have grammaticalized the most. Van de Velde and Lamiroy (2016), for example, argue that the different rates of loss of dative external possessor constructions across Romance and Germanic can be explained by different rates of syntagmatic tightening of NP constituents, which they analyze as an increase in configurationallity of NP constructions.

The three sets of studies briefly mentioned above illuminate different corners of processes of grammatical change in languages related genealogically and/or areally. Each of the studies focuses on different aspects of this phenomenon. Verkerk (2014) concentrates on the typological profile of the encoding of a well defined semantic domain. Enfield (2003) identifies the almost identical pathways of change of a semantically comparable lexical element in areally closely-related languages. Finally, a set of studies has argued that across the Romance and Germanic languages, comparable processes of change seem to advance in different rates in different languages, with some languages tending to be "further along" grammaticalization clines than others.

This chapter complements these studies by concentrating on the crosslinguistic patterns of change in the use of three sets of cognate verbs in Indo-Iranian. That is, instead of holding the semantic or grammatical domain constant, comparing its associated grammar in a set of related languages, the analysis in this chapter hold the cognate lexical material constant, comparing its usage patterns in a set of genealogically related languages.

9.2 The Data

The data used for this chapter is composed of the distribution of the sets of construction types identified in Chapters IV through VI. The verbs chosen for this study are verbs which tend to be involved in or affected by processes of grammatical innovation and are generally more prone to develop some sort of polyfunctional usage. That is, the uses of these verbs tend to extend from their original functions in several distinct functional directions, including the types of events and states they express, with or without changes in the coding means involved, and more grammaticalized functional directions, as in the case of auxiliary construction types.

The selection of verbs that are more prone to polyfunctionality is necessary if one wishes to compare several processes in the aggregate, rather than to concentrate on a single process or to restrict the functional domain involved in the grammatical change. Comparisons of related verbs which are less prone to polyfunctionality (or not prone to it at all) would not yield informative or interesting aggregate results because there would not be enough instances of innovation and change to create a "forest" from the "tree(s)".

There are several motivations for choosing the Indo-Iranian language family for this study. It is large (over three hundred identified languages) and has relatively deep and broad diachronic documentation, and relatively good textual documentation of several contemporary minority languages (despite clear lapses). The identification of major branches of the family is more or less clear, and despite some obvious instances of contact and cross-branch influence (e.g., the status of Persian during the Moghul rule in Northern India), there are clear differences between the grammars of Indo-Aryan and Iranian languages. Furthermore, Indo-Iranian exhibits a great deal of crosslinguistic

grammatical variation in many typological variables (e.g., clause alignment, see Haig 2008 for Western and Central Iranian, Verbeke 2012, Deo & Sharma 2006 for Indo-Aryan).[95]

As mentioned several times above, the construction types identified for each verb differ in their crosslinguistic distribution. A small number of construction types were identified from each of the sixteen languages analyzed here, and many were identified only in one or two languages. Recall that the different construction types are not defined based on constructional cogency or shared innovations, but rather are defined as comparative concepts, following Haspelmath (2010; see also Stassen 1985:14, Croft 2016). Thus, an attestation of the same construction type in two languages does not necessarily entail a shared innovation, since the match might be the result of a shared retention or parallel innovation. In fact, crosslinguistically shared attestations of construction types are more likely to be a result of parallel rather than shared innovation.

Even when a specific construction type can be identified throughout the Indo-Iranian family, and can also be reconstructed as existing in some proto-stage, this does not mean that its wide distribution is a result of a retention of a Proto-Indo-Iranian construction. Across Indo-Iranian, myriad processes of change in argument flagging and verbal indexation have led to innovations in alignment patterns in general, and in argument coding properties in particular. Thus, Kotia Oriya (Eastern Indo-Aryan) constructions with nominative-accusative alignment and Ziyarat Mazandarani (Central

[95] There are other language families, such as Uto-Aztecan or Bantu, in which large enough subfamilies can be discerned, and a good number of cognate verbs or cognate constructions is identifiable in published textual material. Further, the exercise presented here can be repeated in Indo-Iranian by choosing a different set of languages.

Iranian) constructions with nominative-accusative alignment, are not a shared retention of the Proto-Indo-Iranian accusative alignment but are instances of parallel innovation that result in clauses with a nominative-accusative alignment. The differential flagging of P arguments in these constructions, for example, is done by non-cognate markers, innovative in their respective languages: =(*r)ā* in Ziyarat Mazandarani, =*ke* in Kotia Oriya. While the occurrence of the construction type in which "do/make" is accompanied by an A and a P argument can be reconstructed in Proto-Indo-Iranian, the occurrence of this construction type across Indo-Iranian is not (necessarily) the result of retention.

In some sense, it could be argued that all the constructions in which "do/make", "be/become", and "give", are used are likely cognate, since they all have their origin in the constructions of Proto-Indo-Iranian ancestors of these verbs. This does not mean, however, that crosslinguistically shared occurrences of construction types are not the result of parallel innovative processes. Many instances of crosslinguistically attested construction types in which "do/make", "be/become", or "give" function as the main verbal predicate, make use of non-cognate markers in the flagging of different NPs associated with the verb (often those NPs expressing the affected participants). This can be seen in examples (9.1a-c), where "be/become" verbs are used in clauses expressing predicative possession in Heine's 1997 "goal schema". The markers used to flag the possessor in all three constructions (*rāy* in Middle Persian, *ne* in Gujarati, and *ke* in Kotia Oriya) are functionally equivalent but are not cognates. None of these markers existed in Proto-Indo-Iranian or Proto-Indo-Aryan, and they are all innovations in different Indo-Iranian branches.

(9.1a) *ud ōy wirāz rāy haft xwah būd h-ēnd*
and DEM wiraz to seven sister be.PST be.PRS-3PL
'and Wiraz had seven sisters' (Middle Persian, AWN 2.1)

(9.1b) *raja=ne ek kûwəri ch-e*
king=OBJ one princess be-PRS.3SG
'the king had one princess' (Gujarati, Christian 1987:276)

(9.1c) *se dangɽ-i=ke pila oi ga-la*
DEM youngin-F=OBJ child be go-PST.3SG
'The young girl had a child' (Kotia Oriya, Gustafsson 1973a:194)

Moreover, several instances of intralinguistic parallel innovation can also be identified. In some languages, two distinct constructions are classifiable as the same construction type: this can be illustrated by two Middle Persian "be/become" predicative possession constructions. Examples like (9.1a), in which the possessor is flagged by *rāy*, co-exist with examples like (9.2), where the possessor is expressed by the clitic pronoun. The use of the clitic pronoun to express the possessor is a retention of an older construction that is also attested in Old Persian, but an innovation of the Iranian branch. The use of the postposition *rāy* to flag possessors has to be an innovation since the postposition itself is an innovation (see Lazard 1963, for example). Both the construction demonstrated in (9.1a), and the construction in (9.2), are instances of predicative possession with "be/become" in Heine's goal schema (see Chapter V).

(9.2) *ke=šān pad gētīg stōr būd*
REL=3PL in world cattle be.PST.3SG
'(people) Who in the material world had cattle' (Middle Persian, AWN 77.3)

Another common domain in which innovations tend to be parallel rather then shared is N-V complex predication with "do/make". As was shown in Chapter VII, there are multiple distinct pathways of change which could result in structurally equivalent, or at least similar, N-V complex predicate constructions. Parallel innovation in N-V complex predication is sketched again below, but for now, it is demonstrated by examples (9.3a-b) below, from Nagamese and Middle Persian. Both examples are instances of the same construction type, in which "do/make" is accompanied by three nominal elements: an NP encoding the type of event expressed in the clause (the deed element), an A argument, and an oblique object flagged by a comitative marker. In (9.3a), from Nagamese, this oblique object is flagged by *ləgot* 'with, by' (or *logote*, in the orthography of the Nagamese Bible), and in (9.3b), from Middle Persian, this oblique object is flagged by *abāg* 'with'. These two markers are not cognate, and neither can be reconstructed to the last common ancestor of Nagamese and Middle Persian, i.e., Proto Indo-Iranian. Thus, the Nagamese and Middle Persian constructions are likely to have developed separately long after the split between Iranian and Indo-Aryan.

(9.3a) *bak ləgot judo kor-i bak mar-ise.*
tiger COM fight do-PTCP tiger die-PST
'he fought with the tiger and killed it' (Nagamese, Sreedhar 1985:194)

(9.3b) *mard-ēw* *ke* *abāg zan* *ī* *xwēš* *pašn* *kun-ēd*
man-INDEF REL with woman LNK REFL contract do.PRS-3SG
'a man who makes an agreement with his wife' (Middle Persian, RAF Q62)

Finally, some shared innovations are actually attested in the data. Consider, for example (9.1b) in which the possessor is flagged by *=ke*. In Kupia, a sister Eastern Indo-Aryan language, the same construction type is attested, in which the possessor is flagged by *=ka*, a cognate of Kotia Oriya *=ke.*

The data for this chapter, then, consists of the crosslinguistic distribution of construction types. The evolution of these construction types in different languages is the result of different diachronic processes, and an attestation of the same construction types in two or more languages can be the result a shared retention of an old inherited construction, an innovation shared by two (or more) languages, or independent parallel innovations.

9.3 Method

The questions asked in this chapter do not target any specific historical processes in particular. As far as the questions asked in this chapter are concerned, all crosslinguistic similarities are equally important. Any method chosen to analyze the crosslinguistic distribution of construction types, then, must not favor one historical source for similarity, e.g., shared innovation, over another, e.g., parallel innovations.

The Neighbor-Net analysis (Bryant & Moulton 2004) employed for examining the crosslinguistic distribution of different construction types does not assume that the evolution of the data is necessarily tree-like, nor does it assume that similarities or

dissimilarities are the result of a specific historical process. Neighbor-Net and the networks it constructs have often been referred to in the linguistic literature as "phylogenetic", but both the method and its results are not phylogenetic as they do not provide an analysis of the evolution of data based on shared innovations.[96] The Neighbor-Net algorithm considers only similarity between taxa (i.e., of the purposes of this chapter, languages). Thus, two languages could end up clustered together because they share innovations of construction types, share retentions of constructions type, or share parallel innovations of construction types. This method, then, simply provides a good measure of how the different taxa (i.e., languages) cluster together based on their aggregate or overall similarity. The result of applying this method can be profitably interpreted against a known, independently developed family tree, but the results themselves do not provide a phylogenetic analysis. As most of the shared occurrences of construction types in the data analyzed here involve instances of parallel, not shared, innovation, Neighbor-Net analysis with its agnostic treatment of causes of similarity is especially appropriate.

In order to construct the Neighbor-Nets, I first constructed three matrices representing the crosslinguistic presence and absence of construction types across the sixteen languages, one for each of "do/make", "be/become", and "give". The columns of each matrix represent the different construction types attested in the data; the rows of each matrix represent the different languages. Each cell in these matrices is, then, at the intersection of a specific construction type and a specific language. If the construction

[96] The reasons for that, apart from the sociology of the field of historical linguistics, could be that Bryant & Moulton 2004 presented Neighbor-Net in their abstract as a method of constructing "phylogenetic networks". Whether the networks produced by the Neighbor-Net algorithm are phylogenetic depends mostly on whether similarities in the data used to construct them are likely to be the result of genetically shared innovation.

type is attested in data from this specific language, then this cell equals 1, otherwise it equals 0.

As noted earlier in this chapter, the Neighbor-Nets were constructed using SplitsTree4 (Hudson & Bryant 2006), a free open-source software for the construction of different type of networks and trees. The software requires the input of a presence-absence matrix, uses it to calculate the hamming distances between all pairs of taxa and implements the Neighbor-Net algorithm to plot a Neighbor-Net based on these distances. I repeated this procedure using the R Phangorn package (Schliep 2011), which uses Euclidean distances are used.

Finally, despite the many practical and methodological advantages that the Indo-Iranian language family has to offer for such a study, there are some disadvantages that preclude, or at least encumber, the application of some phylogenetic methods such ancestral state reconstruction, which is method which can be used to reconstruct the presence or absence of different grammatical constructions in different ancestral states of a language family. The input this method requires constitutes a fully developed tree of the language family studied and a model of the different possible pathways of change.

Such a model of pathways of change can be constructed based on the analysis of the historical development of certain construction types, such as the analysis provided earlier in Chapter VII, which proposed several pathways leading from lexical or heavy usage of "do/make" to N-V complex predicate constructions. These pathways can be the basis of this model. The problem with applying ancestral state reconstruction to the data analyzed in this chapter, however, is the lack of a fully articulated family tree of either Iranian or Indo-Aryan.

Since Tedesco's 1921 study, the Iranian language family has been considered to be composed of two main branches: Eastern and Western Iranian, with some further split such as the western branch into North-Western and South-Western. The recent emergence of data from modern Iranian languages, especially minority and endangered languages, as well as newly discovered data from extinct languages such as Bactrian and Sogdian, however, has raised many doubts about the validity and accuracy of the major branches in Iranian. Korn (2016) has shown that the East-West division and the subsequent Northwest – Southwest, branching scenario is very problematic in light of this data. She convincingly argues for a ternary, not binary, branching of the Iranian language family into Western, Central, and Eastern branches. Under her analysis Sogdian, Yaghnobi, and Bactrian, which were formerly considered Eastern Iranian languages, together with most Northwestern Iranian languages, would compose the Central Iranian branch. In terms of the number of attested languages, the Central Iranian branch would be by far the largest. In the data used for this dissertation, Gorani, Sivandi, and Ziyarati which would have been considered Northwest Iranian languages, would all belong to the Central Iranian branch, together with Sogdian and Bactrian, which were formerly considered Eastern Iranian. The internal subgrouping of languages of this branch, however, is only partially analyzed by Korn 2016. She shows that Sogdian (and Yaghnobi) and Bactrian branched away from the Central Iranian languages first, but further internal subbranching of the main group of languages remains to be articulated.

The picture of subbranching in the Indo-Aryan family is also far from resolved. Masica (1991:446-463) surveys twelve different proposals for subgroupings of New Indo-Aryan languages. These twelve proposals share some subgroupings but differ with

respect to others. One of the points of tensions is, for example, the exact position of Gujarati in the family tree. Furthermore, none of the hypotheses presented by Masica is concerned with Old and Middle Indo-Aryan languages. The exact classification of Pali, for example, is also still unresolved (see Oberlies 2003, and references therein, for an overview).

Thus, while the major subbranching of Indo-Aryan and Iranian is known to a large degree, the more recent diversification and resultant family tree have not yet been clearly agreed upon. Given the current state of knowledge of the internal branching across Indo-Iranian, the reconstruction of an ancestral states of the usage of "do/make", "be/become", and "give" would be problematic.

9.4 Is There a Historical Signal in Overall Usage of Cognate Verbs?

This section presents the Neighbor-Nets constructed using the process described in Section 9.3 above, interpreting them in light of the known subbranching of Indo-Iranian. I will argue here that overall, there is a correlation between known genealogical subfamilies of Indo-Iranian and the patterns in which "do/make", "be/become" and "give" are used crosslinguistically. The three Neighbor-Nets differ in the genealogical scope of clusters identified in each. The crosslinguistic distribution of construction types of "do/make" and "give" in the Iranian languages analyzed here is reliably different from their Indo-Aryan counterparts. There is, however, little evidence for heightened similarities in the crosslinguistic distribution of construction types between members of more recently evolved subfamilies of Iranian or Indo-Aryan. The opposite is found for "be/become": here, there is little evidence for a difference in the level of the major

branches of Indo-Iranian, but there are clear clusters composed of languages of tighter historical relationship.

The Neighbor-Nets for “do/make” and “give” are shown in Figures 9.1 and 9.2 below. The overall pattern illustrated in both networks is the same. The thick solid black lines capture the nodes representing the Iranian languages in both figures (plus Hindi in Figure 9.2 illustrating the Neighbor-Net for “give”). With the exception of Bactrian, the nodes representing all the Iranian languages in the sample are clustered together on one side of the network. In both networks, there is little to no evidence for clusters which include languages belonging to specific subbranches of Iranian. For “do/make”, Gorani and Sivandi, both Modern Central Iranian languages, are close to each other but the other modern Central Iranian language, Ziyarat Mazandarani, is located on almost exactly the opposite side of the Iranian cluster, by the node representing Sogdian, a Middle Central Iranian language. A similar picture emerges for “give”. Finally, in both networks, Bactrian, a Middle Central Iranian language, is not clustered together with the other Iranian languages.

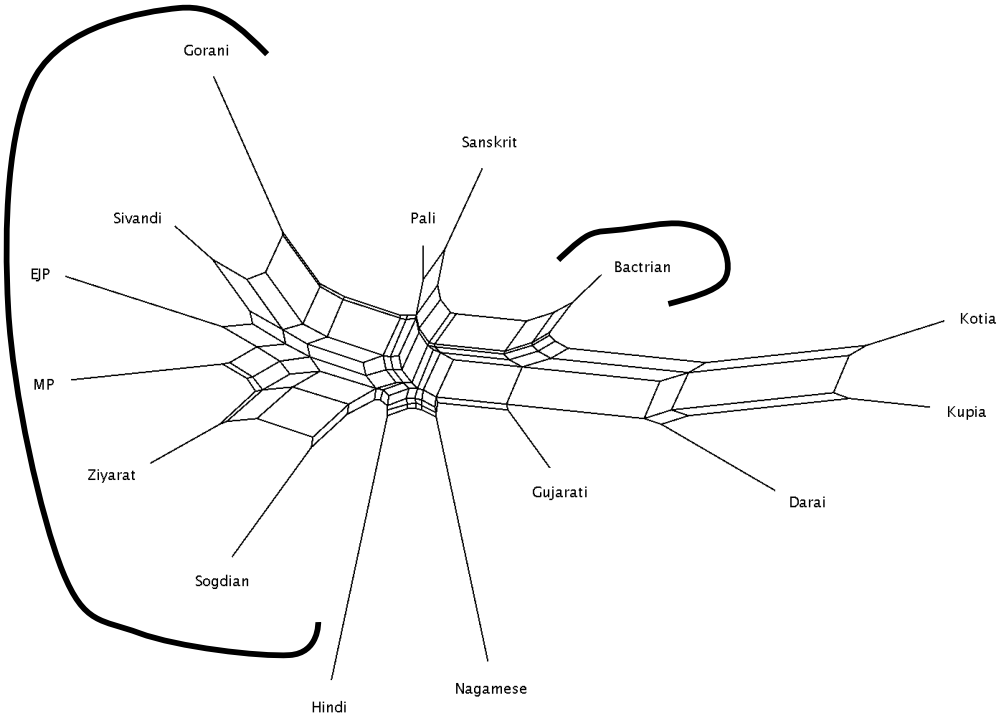


Figure 9.1: Neighbor-Net of "do/make"

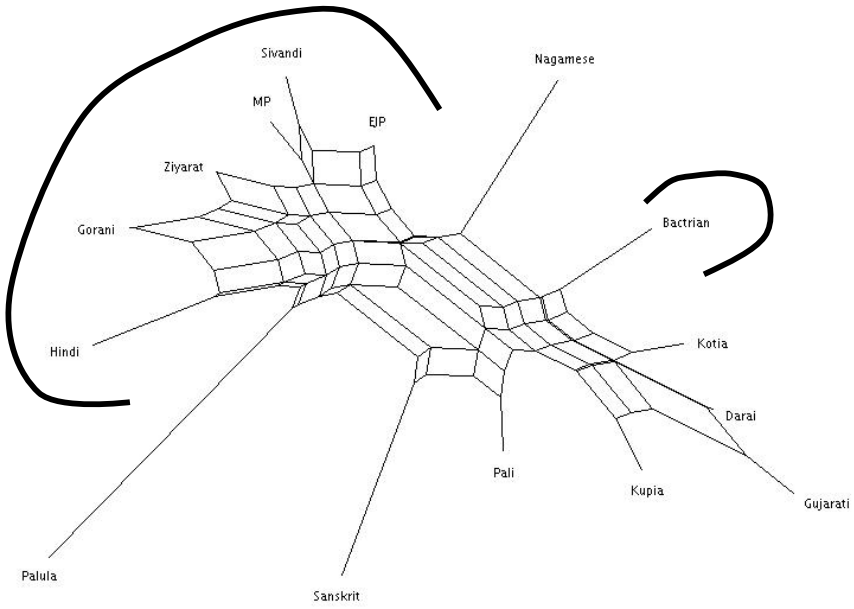


Figure 9.2: Neighbor-Net of "give"

Unlike the tighter cluster of Iranian languages in Figures 9.1 and 9.2, the Indo-Aryan languages of the sample are more spread apart. This suggests that the usage of "do/make" and "give" across Indo-Aryan is more varied than in Iranian, that Indo-Aryan languages have a smaller number of shared occurrences and shared absences of the different construction types attested for the two verbs. This increased variation, however, is accompanied by at least one instance of a clear cluster of genealogically closer Indo-Aryan languages. In the Neighbor-Nets for "do/make" and "give", Kotia Oriya, Kupia, and Darai, all Eastern Indo-Aryan languages, are clustered more closely together. Nagamese, however, is never clearly clustered together with these three languages. In the network for "give", Gujarati, which does not belong to the Eastern Indo-Aryan branch, and is spoken in the western coast of India, is clustered more closely with to the Eastern Indo-Aryan languages. The colloquial Gujarati variety analyzed here, however, lost all construction types in which cognates of "give" function as the main verbal predicate (these were replaced by *āpī-* "give").

In both networks, Sanskrit and Pali, an Old Indo-Aryan language and a Middle Indo-Aryan language respectively, are clustered closer together. This suggests that the overall usage patterns "do/make" and "give" in the older Indo-Aryan language is reliably distinct from their usage in the modern languages. In contrast, Middle Persian and Sogdian, two Middle Iranian languages, are clustered with the other Iranian languages with "do/make", and Middle Persian is clustered with the other New Iranian languages in the "give" network (Sogdian lost all non-derived forms of the Proto-Iranian **daH,* and "give" is expressed in Buddhist Sogdian by a combination which roughly translates as "bring forth"). This suggests that main innovative processes which distinguish Iranian

"do/make" and "give" from the Indo-Aryan "do/make" and "give" are already well on their way during the Middle Iranian stage, with the exception of Bactrian.

The diversification pattern identified with "do/make" and "give" is not the result of branch specific construction types which occur in all languages of one branch and are absent from the other. That is, for every construction type identified in the data, it is either found in at least one language from each branch, or is identified only in a partial set of the languages of one branch. For example, one very Indo-Aryan construction type is the V-V compound verb construction. It is found to varying degrees across all New Indo-Aryan languages (see Hook 1975, 1991 for details), and as Masica (1991:326; see also Slade 2013) notes, it is an innovation of New Indo-Aryan languages. Indeed, in the Sanskrit and Pali data analyzed here, there are no occurrences of such construction type. Thus, this construction type cannot be the (sole) contributor toward the division between Iranian and Indo-Aryan languages with "give".

Finally, in both networks, Hindi is the Indo-Aryan language closest to the Iranian cluster. In the network for "give", Hindi is included within the cluster of Iranian languages. This situation is analyzed in slightly more detail below. For now, suffice it to mention that the position of Hindi could be motivated by the intense contact it had with Persian during the Moghul rule of Northern India (sixteenth and seventeenth centuries).

In contrast to the networks illustrating the crosslinguistic distribution of construction types of "do/make" and "give", the network illustrating the diversification of "be/become" cannot be argued to have a clear Indo-Aryan vs. Iranian split. The Neighbor-Net constructed for "be/become" is plotted in Figure 9.3 below.

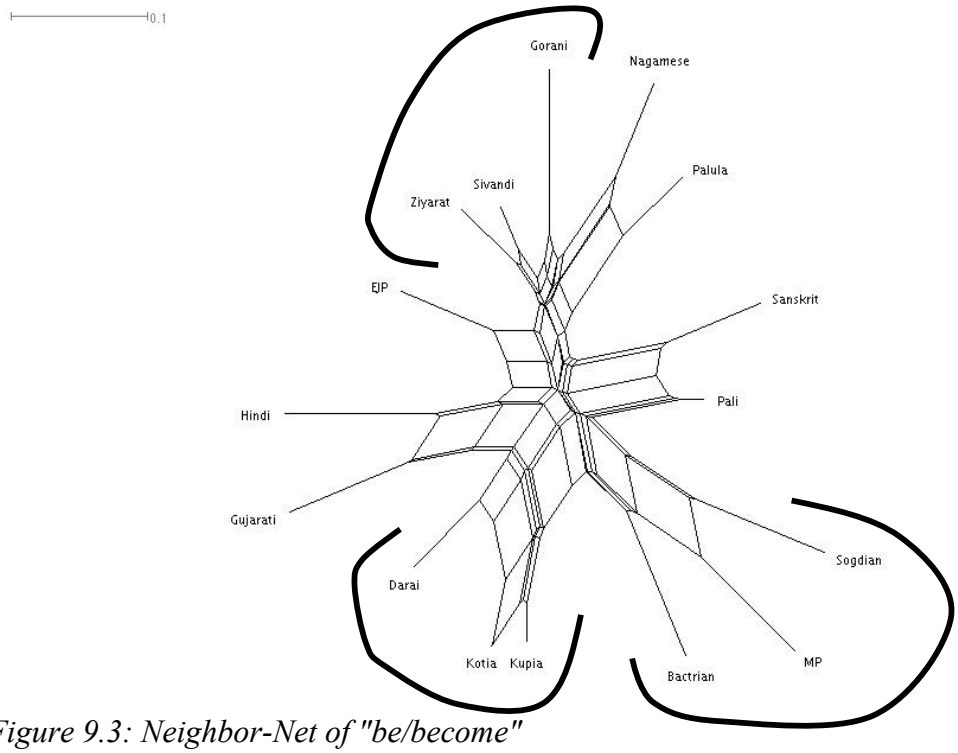


Figure 9.3: Neighbor-Net of "be/become"

The clustering of nodes representing the overall usage of "be/become" across the Indo-Iranian languages does not seem to correlate with the major branching of Indo-Iranian into Indo-Aryan and Iranian. Instead, there are several smaller clusters which include one or two languages with some genealogical or diachronic relationship between them. These clusters are marked in Figure (9.3) with a thick black line.

Two of the three clusters include modern languages which belong to a single subbranch of Indo-Iranian. One such cluster, located in the upper part of the Neighbor-Net, consists of the three modern central Iranian languages analyzed here: Ziyarat Mazandarani, Sivandi, and Gorani. The second cluster of this type, on the lower left hand side of the Neighbor-Net and consists of Darai, Kotia Oriya, and Kupia, all eastern Indo-Aryan languages. The third cluster, consists of languages which belong to the Middle

Iranian historical stage, and is found adjacent to the Eastern Indo-Aryan languages. This cluster includes Sogdian, Middle Persian, and Bactrian.

At this point, it should be clear that languages from the same branch or subbranch of Indo-Iranian are more similar in their aggregate usage patterns of "do/make", "be/become", and "give" then across these branches or subbranches. In Figures (9.1) and (9.2), for example, the Iranian languages are clearly clustered together. In Figure (9.3), there are clear clusters composed of languages which belong to the same subbranch of Indo-Iranian, such as the one composed of the three Modern Central Iranian languages, and the one composted of the Eastern Indo-Aryan languages.

The three verbs which usage patterns are analyzed here, then, differ in the scope of genealogical clusters identified. With "do/make" and "give", the scope of the major clusters of languages consists of the Iranian languages of the sample with the exclusion of Bactrian, and the inclusion of Hindi with respect to "give". There are some further, smaller clusters for both these verbs, which are composed of three of the Eastern Indo-Aryan languages analyzed here (Kupia, Kotia Oriya, and Darai). For "be/become", there are no clusters corresponding to one of the major branches of Indo-Iranian. Instead, smaller clusters can be identified composed of genealogically and diachronically related language. These clusters have scope over New Central Iranian languages, New Eastern Indo-Aryan languages, and Middle Iranian languages.

9.5 The Historical Processes behind Diversification Patterns

Now that the first two questions posed earlier in this chapter have been answered, a hypothesis can be offered as an answer to the third question, i.e., what drives the difference in the type of clusters found with "do/make" and "give" vs. "be/become. In this section, I will argue that the difference is the result of the varying types of pathways of grammatical change which gave raise to the key construction types identified with "do/make" (as representative of those found with "do/make" and "give") and "be/become".[97]

The pathways involved in the evolution of the more "Iranian" construction types of "do/make" involve several dependent stages, and consist of changes in participant coding that were motivated by analogy to coding options available to speakers and used elsewhere by them. Speakers for whom no such analogical source is available are less likely to develop these construction types as an innovation. In contrast, I argue that many pathways along which different construction types of "be/become" have developed involve crosslinguistically common and accessible metaphors and metonymies, and do not involve changes in the overt coding of any component of the construction. These changes are more or less equally likely across the different languages.

A look at the data behind the Neighbor-Net in Figure 9.1 reveals that most of the construction types driving the clustering of Iranian languages, that is, construction types which are more commonly found in Iranian languages, are those which are often involved in N-V complex predication. That is, these are construction types in which "do/make" is

[97] I concentrate here on construction types found with "do/make" as they illustrate the different pathways of change more clearly. A similar argument, however, can be made for "give".

accompanied by a deed element which expresses the main propositional content of the clause, and one to three other NPs or PPs expressing the different participants in the event. In Chapter VII, it was argued that the evolution of different types of N-V complex predication in Indo-Iranian occurred, and is still occurring, along several distinct and partially overlapping pathways. The source constructions involved in all these pathways are one of the "heavy" or lexical uses, of "do/make", in which the verb expresses the creation of a tangible object or the caused change of state of a referent. These two construction types have been described and analyzed in Chapter IV, and are demonstrated in (9.4 a-b).

(9.4a) *naxud.pelu* *kārd-en*
pea.cooked.rice do.PST-3PL
'they (the women) made rice with peas' (Ziyarati, Shokri et al. 2013:82)

(9.4b) *amalõ=ko* *kʰuʃ* *kar-o*
workers=OBJ happy do.PRS-3SG
'he made the workers happy' (Hindi, Premchand 2017[1936]: 3.22)

The different pathways identified in Chapter VII all involve several stages which at some point involve changes in the overt coding properties of at least one constituent in the construction.[98] Table 9.1, repeated from Chapter VII, summarizes one of the pathways identified there, and its starting point is clauses like (9.4b) above that express caused change of state, and its end point is N-V complex predicate constructions where

[98] As argued in chapter VII, it is likely that changes in behavioral properties, such as availability for certain derivational properties and relativization occur earlier in the process.

the deed argument is coded like a P argument, at least in terms of argument indexation on the verb in ergative constructions. Recall that the first four stages in this pathway do not involve any overt changes in coding properties in any component of the clause. The NP expressing the affected participant is coded as a P argument in stages 1 through 4. In all the languages analyzed here, at least three of the first four stages can be identified. In Chapter VII, each of these stages was demonstrated by examples from several languages, most notably Middle Persian.

Table 9.1: the evolution of N-V complex predicates.

Stage	Schematic structure	Schematic example
Stage (1)	$NP_{[A\ argument]}$ $NP_{[P\ argument]}$ ADJ "do/make"	She made him happy
Stage (2)	$NP_{[A\ argument]}$ $NP_{[P\ argument]}$ $NP_{[state\ semantics]}$ "do/make"	She made him shelter
Stage (3)	$NP_{[A\ argument]}$ $NP_{[P\ argument]}$ $NP_{[transitive\ event]}$ "do/make"	She made him punishment
Stage (4)	$NP_{[A\ argument]}$ $NP_{[P\ argument]}$ $NP_{[less\ transitive\ event]}$ "do/make"	She made him advice
Stage (5)	$NP_{[A\ argument]}$ $OBL_{[oblique\ argument]}$ $NP_{[less\ transitive\ event]}$ "do/make"	She made advice to him
Stage (6)	$NP_{[A\ argument]}$ $OBL_{[oblique\ argument]}$ $NP_{[event;\ P\ argument]}$ "do/make"	She made trust on him

The shift from stage 4 to stage 5 seems to be where "do/make" cognate verbs vary between Iranian and Indo-Aryan languages. This shift involves a reanalysis of the affected participant as not a prototypical affected participant in transitive events, and an actualization of this reanalysis by the flagging of the affected participant. The actualization is a change in the coding properties of the affected participant from a P argument to an oblique object flagged by an adposition. As this shift has been discussed in chapter VII, it will be only briefly repeated here with two examples from Sivandi. In (9.5a) the main verb is "ask", a lexical heavy question verb. The addressee of the

question is flagged by *ba* ‘to, for’. The clause in (9.5b) also expresses an event of asking a question, but with a N-V complex predicate construction with *sohāl* ‘question’ as its nominal component and “do/make” as its verbal component. Here too, the addressee of the question is flagged by *ba* ‘to, for’.

(9.5a) *ī* *ma'nā=rā* *ki* *ba te* *pors-iy=eš* *en?*
DEM meaning=OBJ who to 2SG ask.PRS-3SG=3SG be.PRS.3SG
‘who asked you the meaning of this?’ (Sivandi: Lecoq 1979:113)

(9.5b) *Ye* *rū=i* *ba mehtar sohāl=em* *kerd*
one day=INDEF to groom question=1SG do.PST
‘one day, I asked the groom’ (Sivandi, Lecoq 1979:113)

The second shift in this pathway that involves changes in the coding of participants is from stage 5 to stage 6. This shift involves a reanalysis of the syntactic status of the deed element as a P argument, since it is only NP in the clause apart from the NP expressing the doer participant, and an actualization of this reanalysis through overt coding of the deed element as a P argument.

The other pathways leading to the different N-V complex predication constructions outlined in Chapter VII all involve, in one way or another, changes in the coding of the deed element, the affected participant, or both. These changes in coding are more likely if there is some existing construction, or a part of a construction, available to speakers which can serve as an analogical source for extension.

Many of the pathways by which N-V complex predication arises, then, involve multiple stages and actualization of reanalysis involving changes in the coding of the

affected participant and the deed element. These actualization processes tend to occur by means of analogy to some pre-existing coding option available for speakers. Most of these involve the availability of oblique objects, which are more common in Iranian than in Indo-Aryan languages.

In contrast to these processes, many of the processes which give rise to construction types of "be/become" are abrupt, being composed of only one or two stages, and do not involve overt changes in the coding of NPs associated with "be/become". The starting point of these changes is construction types which express some well defined nominal predication domain(s), such as core nominal predication or predicate locative or existential. The changes involved include an extension of the semantic type of NP used in a particular function in these construction types, which changes the orientation of the entire clause to express a different nominal predication domain. Importantly, there are no changed in the grammatical coding of any component of these construction types at any time.

The fact that nominal predication domains are often co-expressed by the same configuration of structural coding means has been at the primary focus of Chapter VIII. There, it was shown that across Indo-Iranian, distinct nominal predication domains are often co-expressed by clauses with the same copular element, the same flagging of non-verbal constituents, and the same relative word order. Examples (9.6a-b), repeated from Chapter VIII, demonstrate one such co-expression pattern. In both examples, "be/become" is accompanied by two unflagged NPs. The clause in (9.6a), however expresses core nominal predication and the clause in (9.6b) expresses predicate location.

(9.6a) *īrij kuř=aš biya*
irij son=3SG be.PST.3SG
'Irij was his son' (Gorani, Mahmoudveysi et al. 2012, 5:8)

(9.6b) *usā āsā faransa biya*
master then France be.PST.3SG
'At that time, the master was in France' (Gorani, Mahmoudveysi et al. 2012, 4:65)

The combination of two unflagged NPs and "be/become" is usually used, across Indo-Iranian, to express core nominal predication. That is, this configuration of structural coding means is usually found in construction types expressing equation, predicate property, or proper inclusion. In such clauses, the predicate is usually a noun or an adjective expressing a property, a referent, or a set of referents. The shift in the nominal predication domain from predicate property to predicate location is the result of a change in, or extension of, the semantics of the predicate noun to express location names such as "France" in (9.6b). The change involves only the type of nominal elements in the function of the predicate in these clauses, and the structural coding means (such as flagging) remain unchanged.

A similar situation is shown in (9.7a-b) below, from Darai. In both examples, "be/become" is accompanied by two unflagged NPs. In (9.7a), the predicate is *ramrə* 'good' and the clause expresses core nominal predication. In (9.7b), on the other hand, the predicate is *cʰawa* 'child' and the clause expresses predicative possession: the clause expresses not that his wife was not a son (despite this being a true fact), but that she had no son, i.e., that she hadn't given birth to a child.

(9.7a) *te=rə* *kərmə ramrə nidzə bʰəi-lə*
2SG.OBL=GEN fortune good NEG be-PST
'your fortune was not good' (Darai, Dhakal 2013:107)

(9.7b) *ukʰ=rə* *dulhi=k* *cʰawa nidz=bʰəi-lə*
3SG=GEN wife=3SG.POSS son NEG=be-PST
'his wife did not have a son (lit. his wife was not a son)' (Darai, Dhakal 2013:79)

Changes in the nominal predication domain expressed by a clause, without changes in the coding properties of any of its constituents, are not limited to construction types illustrated by (9.6 – 9.7). Similar changes can be found when configurations of coding means which express predicate locative or existence are co-opted to express core nominal predication, especially predicate property and proper inclusion.

Such a shift can happen in two scenarios. In the first scenario, illustrated by (9.8a), the phrase is headed by the postposition *madʒí* 'in', expressing not the location of the dragon, but its state. Thus, the clause does not express predicate locative, but rather expresses predicate property. In (9.8b), the opposite type of extension occurs. The figure NP expresses the property or state of being mysterious, and the locative phrase expresses the issue which has this property. Again, there are no changes in the coding properties of any of the constituents of these clauses. The only changes are in the nominal predication domain these clauses express.

(9.8a) *níndram madʒí hín-u*
sleep in be.PRS-MSG
'(The dragon) was in sleep' (Palula, Liljegren & Haider 2015:78)

(9.8b) *ī* *ye* *serr-i* *dar kār-e* *h-and*
DEM INDEF secret-INDEF in deed-DEF be.PRS-3SG
'there's an enigma in this affair' (Sivandi, Lecoq 1979:128)

Changes like those illustrated in (9.6 – 9.8) involve only changes in the semantic field, or the type, of one or two of the clause constituents. They do not involve any change in structural coding means, particularly the flagging of constituents, and hence they do not depend on the coding means available to speakers for analogy. Moreover, it appears that these changes are based on crosslinguistically common and available metaphors and metonymies equating states with location, possessors with locations etc. Thus, these changes are about equally likely across branches of Indo-Iranian, and it does not seem that any one branch has an advantage over the other in respect to the availability of those metaphors and metonymies.

The roughly equal likelihood of these changes, together with the fact that they do not seem to depend on pre-existing argument coding strategies, means that they occur sporadically across Indo-Iranian, and across the data analyzed here. The result is that the somewhat accidental crosslinguistic distribution of these construction types blurs the similarities that would arise as a result of diachronically determined factors. This leads to the hypothesis that if we ignore construction types which arise as a result of the processes mentioned above, the similarities in the usage of "be/become" between genealogically closer languages would increase.

To quickly test this hypothesis, I constructed another Neighbor-Net based on the distribution of "be/become" across the different construction types, but this time I ignored all construction types which arose by the processes just described. The result is

the Neighbor-Net in Figure (9.4) below. Clusters of Indo-Aryan languages are marked by the two thick black lines in the figure. This network still does not show the strong split between Indo-Aryan and Iranian in the the networks in Figures (9.1) and (9.2). It is, however, much closer to such a clustering then the network in (9.3), which took into account all construction types with "be/become". Specifically, in network (9.4) below, the Indo-Aryan languages cluster more tightly with each other while the Iranian languages are more widespread. Furthermore, two Indo-Aryan languages, Palula and Nagamese, are clustered together with the three Modern Central Iranian languages, Sivandi, Gorani, and Ziyarati. The overall picture that emerges from figure 9.4, however, is tidier than the one which emerged in figure 9.3.

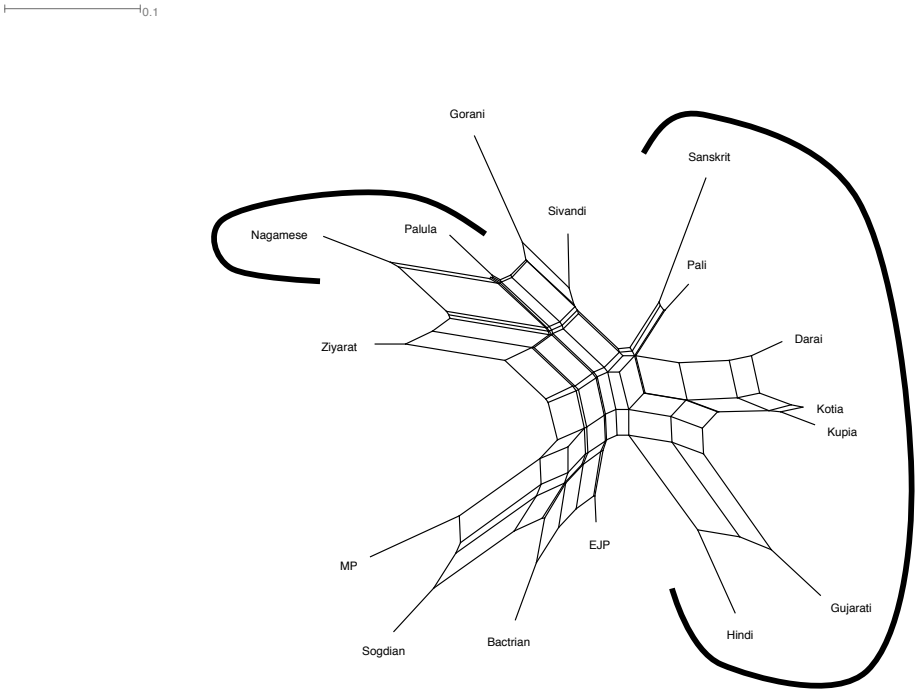


Figure 9.4: Partial "be/become" Neighbor-Net

Section 9.4 showed that in the aggregate, “do/make”, “give”, and “be/become” are used more similarly in closely related languages than in more distantly related languages, and that for “do/make” and “give”, the crucial genealogical units are the major subbranches of Indo-Iranian, Indo-Aryan and Iranian. The behavior of “do/make” and “give” is more similar between Iranian languages, than between Iranian and Indo-Aryan languages. This increased similarity results in the clusters of Iranian languages that are shown in Figures 9.1 and 9.2, and the looser relationship among the different Indo-Aryan languages. The crosslinguistic distribution of construction types with “be/become”, one the other hand, was less sensitive to the major split of Indo-Iranian, and more sensitive to more recent splits which resulted in several smaller diachronically motivated clusters.

This section argued that the different types of clusters found with “do/make” on the one hand, and with “be/become” on the other, are the result of the different types of processes which gave raise to innovative construction types with both verbs. It has been shown here that the construction types which drive the clustering of Iranian languages with “do/make” have evolved in several successive stages which involve innovative flagging of the affected participant as an oblique object. This flagging arose by analogy to flagging of other oblique objects, already used elsewhere by speakers. With “be/become”, on the other hand, several construction types arose by abrupt changes which did not include changes in the coding properties of any constituent. Eight out of the twenty-eight construction types of “be/become” are the result of such changes. The crosslinguistic distribution of these construction types is rather random, and hence blurs many affects which might lead to more historically motivated clusters, such as those found with “do/make”.

9.6 Summary and Further Hypotheses

This chapter investigated the crosslinguistic distribution of construction types. First, this chapter asked whether the crosslinguistic distribution of construction types is correlated with genealogical relations between languages. The answer to this question was shown to be affirmative. It was then determined that the crucial genealogical unit for these similarities differ across the three verbs which usage is analyzed here. The crucial unit for "do/make" and "give" is the major subbranches of Indo-Iranian, Iranian and Indo-Aryan. The crucial units for "be/become" are smaller, more recent subbranches. These were shown in Figures 9.1 – 9.3.

It was then hypothesized that the difference between "do/make" and "give" on the one hand, and "be/become" on the other, has to do with the different types of evolutionary pathways these verbs were involved in. It is difficult to argue for this as a general reason based on the one case study presented here, but it could be hypothesized that the reanalysis and actualization motivated (in part) by analogy to existing constructions leads to similarity in usage patterns within a genealogical unit (i.e., branch), while extensions motivated by metaphor and metonymy blurs such distinctions.

CHAPTER X

CONCLUSIONS AND SUMMARY

This dissertation has studied patterns of functional and structural diversification of three sets of cognate verbs in the Indo-Iranian language family. The verbs chosen for this study, "do/make", "be/become", and "give", are verbs that are crosslinguistically prone to polyfunctional usage. Hence, differences in their usage across the family are to be expected. In order to develop a snapshot of their uses across the Indo-Iranian family, I collected a series of translated naturalistic texts in sixteen Old, Middle, and New Indo-Iranian languages. I extracted all of the tokens of cognate "do/make", "be/become", and "give" verbs from these texts, consisting of 150 to 500 tokens per verb per language. Based on these tokens, I analyzed the patterns in which each of these verbs is used across the family. These analyses were then used in several studies, comparing specific uses of these verbs and the aggregate patterns in which these verbs occur.

Chapter I introduced the study and the research questions. It presented the theoretical orientation of the study, which is a general, non-formal version of construction grammar, and more specifically that of Radical Construction Grammar (Croft 2001). The chapter then motivated the treatment of constructions, both clause-level construction and morpheme-sized constructions, as language specific grammatical entities. Chapter I presented notion of "comparative concepts" (Hockett 1955, Stassen 1985, Haspelmath 2010, Croft 2016). The chapter suggested defining construction types as bundles of "hybrid" comparative concepts, including both functional and crosslinguistically comparable morphosyntactic components. This enables the identification of construction types across the Indo-Iranian language family, and thus provides a basis for a

crosslinguistic comparison of the usage patterns of cognate verbs. Then, Chapter I introduced the main mechanism of language change assumed by this study, following Harris & Campbell 1995: reanalysis (and actualization), extension, and borrowing. As these mechanisms were not originally defined for a constructional model of grammar, the chapter also presented one way in which these mechanisms have been integrated into such a model (Barðdal and Eythórsson 2012, Barðdal & Gildea 2015). Chapter I ended with a presentation of the main structural coding means used in Indo-Iranian languages to code core arguments and oblique objects: relative word order, argument indexation on the verb, and argument flagging.

Chapter II provided a more in-depth introduction to the data used in this study. Since grammatical description (or dictionaries) seldom enumerate all the uses of a verbal lexeme, this study used published naturalistic texts as its main source for data. These are texts collected and published by field workers, prose texts published by native speakers for a native speaking target audience, and critical editions of texts in extinct languages. Texts from living languages are all accompanied by a translation done by a native speaker or in collaboration with a native speaker. Texts in extinct languages are accompanied by translations done by philologists as part of the publication of the critical edition.

Chapters III through VI presented an analysis of the usage patterns in which sets of cognate "do/make", "be/become", and "give" verbs are attested in the data. Chapter III serves as an introduction, restating and illustrating the method of using construction types, bundles of hybrid comparative concepts, for the comparison of the uses of these cognate verbs. Then, Chapter III showed that many construction types identified in the

data are attested in only a single language, and that most construction types are attested in three to ten languages.

Chapters IV, V, and VI, described and analyzed the different attested uses of each verb in the data. Altogether, these chapters identify over sixty distinct construction types for the three verbs combined, and thus confirm the polyfunctional nature of their usage. "Do/make", "be/become", and "give" were found functioning as "heavy" or "lexical" verbs, found in distinct complex predicate constructions, in nominal predication constructions, in serial verb constructions, and in several types of auxiliary constructions. The description of the different construction types identified in the data ranges from short, almost telegraphic descriptions accompanied by one or two examples, to more in-depth analyses of some more challenging uses. Some of analyses presented in these chapters can be used as a basis for a more delicate comparative study of some functional and structural domains, some of which are mentioned below.

Chapters VII and VIII analyze specific patterns in which "do/make" and "be/become" are used in the data. Chapter VII analyzes the use of "do/make" cognate verbs in N-V complex predicate constructions, and Chapter VIII analyzes the use of "be/become" alongside other verbs and copulas in clauses expressing the six nominal predication functional domains. Chapter IX compares the aggregate usage patterns of "do/make", "be/become", and "give" across the sixteen languages analyzed in this study, testing whether these patterns are more similar in languages which belong to the same branch than across branches.

10.1 Some Main Findings

This section describes some of the main findings of this dissertation, focusing on Chapters VII through IX. Some interesting findings, however, are also presented in Chapters IV through VI, but these relate mostly to a specific use (or a related set of uses) of one of the three verbs analyzed in this study.

Chapter VII has two main findings. First, after presenting a taxonomy of N-V complex predicate constructions with "do/make" into five distinct types (following the findings of Mohanan 1997 and Haig 2002 for Hindi and Kurmanji respectively), this chapter showed that these types differ in the range of different situations they are used to express. These differences lead Mohanan and Haig to admit that it is difficult to predict which type of N-V complex predicate would be used to express a specific type of situation. Chapter VII showed, however, that the distribution of N-V complex predicate types is very similar to the distribution of argument frames of lexical or "heavy" verbs identified by Haspelmath (2015) and Lazard (1994, 2002): one type of N-V complex predicate, in which the affected participant is coded as a P argument, is privileged, and is used to express many different types of situations. Other N-V complex predicates, in which the affected participant is coded like an oblique object or a complement clause, are used to express a narrower range of situation types.

Then, Chapter VII presented several distinct diachronic pathways from lexical, or "heavy" uses of "do/make" to uses of this verb in all five types of N-V complex predicates. These pathways were illustrated using data from Middle Persian, the oldest language in which I could identify all five N-V complex predication types with confidence. The illustration of these pathways with data from Middle Persian also serves

to show that despite the justified doubts regarding the existence of N-V complex predicates in Middle Persian (e.g., Korn 2013, who shows that the examples previously used to argue for the existence of N-V complex predicates in Middle Persian are problematic), these constructions are well attested and fully developed at least by Late Middle Persian.

Chapter VIII analyzed the co-expression patterns of the six nominal predication functional domains. It presented a novel method, based on bipartite graphs (a tool often used in social network research), for visualizing and measuring the degree to which two distinct nominal predication functional domains are co-expressed by the same configurations of structural coding means. That is, Chapter VIII proposed a continuous measure, based on the degree to which the sets of structural coding means used to express two distinct domains overlap. This is in contrast to studies such as Stassen 2013a, who focuses on the identify of the copula alone, and Clark 1978 who considers other coding means as well, but never as an ensemble. Chapter VIII then used this method to measure the co-expression degree of (a) core nominal predication (equation, predicate property, proper inclusion) and the predicate locative (similarly to Stassen 2013a), (b) predicative possession and the predicate locative / existential functional domains, and (c) the predicate locative and the existential functional domains.

This chapter showed that even when considering entire configuration of structural coding means, the six nominal predication functional domains are co-expressed, which points to active processes of extension (metaphoric, metonymic) across Indo-Iranian. It further showed that the degree to which core nominal predication and the predicate locative are co-expressed varies across the family, which points to a varying degree to

which the privileged relationship between states and locations (e.g., Jackendoff 1983, 2002, DeLancey 2000) effects the grammar of nominal predication across the family. It then showed that predicative possession and the predicate locative are seldom co-expressed, but predicative possession and the existential are often expressed by the same configurations of structural coding means. Thus, the privileged relationship between possessors and locations also has varying effects on the grammar of nominal predication across the family, but it seldom leads to a co-expression of the predicate locative and predicative possession. Finally, Chapter VIII showed that the Predicate locative and the existential seldom share the same configuration of structural coding means, and often differ in terms of relative word order, thus reaffirming at least one of the findings of Clark 1978.

The method presented in this chapter could potentially be applied in other onomasiological studies, or expanded to other language families.

Chapter IX showed that the aggregate uses of "do/make", "be/become", and "give" are often more similar within branches of Indo-Iranian than across branches. The aggregate uses of "do/make" and "give" are more similar between languages which belong to one of the two main branches of Indo-Iranian, Iranian and Indo-Aryan, than between languages which belong to different main branches. This pattern is not found with "be/become", where similarity is found between languages groups which belong to recently emerging branches (e.g., New Central Iranian or New Eastern Indo-Aryan) or to historical stages (e.g., Middle Iranian).

Chapter IX argued that the difference in the aggregate distribution of "do/make" and "give" on the one hand, and "be/become" on the other hand, is a result of the

different processes of change attested with these verbs. Many changes with "do/make", for example, such as those identified in Chapter VII, involve analogically motivated changes in the coding of participants from P arguments to oblique objects. Such changes are more likely if speakers already use oblique objects elsewhere in the language (or in another language they often use). As the sources of this analogy are more common in Iranian than in Indo-Aryan, they can lead to the differences in uses of "do/make" (and "give") across the two main branches of Indo-Iranian. In contrast, changes in the usage of "be/become" tend to involve only extension, without many changes in the coding properties of constituents. These changes result in the different patterns of co-expression identified in Chapter VIII.

10.2 Some Future Directions

Just as this dissertation provided some answer to questions, it raised others. In this section, I wish to briefly present three such questions which might be of interest beyond the comparative study of the Indo-Iranian language family. These questions involve the diversification processes in the expression of predicative possession with "be/become" and by other means, a more detailed analysis of oblique objects across the family, and some specific uses identified in the data.

Chapter V showed that the expression of predicative possession in Indo-Iranian with cognate "be/become" verbs is a typologically diverse field, and includes all of the schemata proposed in Heine (1997) and some other construction types which do not neatly fit in those schemata. The rise of this variation is tightly connected to the loss of the old Indo-Iranian morphological case system, and the replacement of the old Genitive,

or Genitive-Dative, morphological case forms by innovative adpositions. As a quick glance at the examples in Chapter V would show, however, that this crosslinguistic variation is also accompanied by intralinguistic variation: in many languages, from Middle Persian to Palula, predicative possession is expressed by several distinct constructions with "be/become". Further, in many languages (especially in the Iranian branch), predicative possession is also expressed by a "have" type verb. This situation can be a fertile ground for the study of the process by which competing constructions emerge and their semantic niches are curved (or not) by speakers. The fact that this situation, of a pressure to innovate constructions for predicative possession, is identifiable in *all* Indo-Iranian languages from the tenth century to today, makes this an interesting area for testing various theories of morphosyntactic change.

The gradual loss of the old Indo-Iranian morphological case system has also lead to changes in the strategies different languages use to code, especially flag, participants in different types of situations. In many languages this lead to the innovation of different types of oblique objects, often flagged by different adpositions. Again, as the loss of the old case system affected *all* Indo-Iranian languages from the tenth century (if not prior to that) until today, a comparison of the innovative processes in which new types of oblique objects increase in frequency can be an interesting ground for the testing of different theories of syntactic change.

Finally, Chapter IV through VI identified several construction types in which "do/make", "be/become", and "give" function as auxiliary verbs. In most of these construction types, the function of the auxiliary verb is to express different TAM categories. Beyond the languages analyzed here, one can also find cognate "be/become"

verbs used as auxiliaries in construction expressing evidentiality, especially in Iranian. The function of at least one construction type, however, in which "do/make" functions as an auxiliary verb, seems to tightly interact with the function of the clause in discourse. Chapter IV provided some analysis of the function of this construction type, found in Kupia and Kotia Oriya, but a more detailed analysis, which would also include an analysis of the interaction of this auxiliary construction with compound verb constructions (such as the one described for "give" in chapter VI), might yield interesting results about the exact system of expressing discourse functions in Kupia and Kotia Oriya.

APPENDIX

LIST OF ABBREVIATIONS

ABS	absolute
ABST	abstract
ACC	accusative
ADJ	adjectivizer
AOR	aorist
CAUS	causative
CLF	classifier
CMPR	comparative
COND	conditional
COM	comitative
COMP	complement
CV	converb
DAT	dative
DEF	definite
DEM	demonstrative
DIR	direct
DOM	differential object marker
DRCT	directional
DS	different subject
EMPH	emphatic
ERG	ergative
EXCL	exclamative
F	feminine
FOC	focus
FUT	future
GEN	genitive
GER	gerund
IND	indicative
IMPF	imperfect
IMPR	imperative
INF	infinitive
INST	instrumental
INDEF	indefinite
IRR	irrealis
LNK	linker (ezafe linker)
LOC	locative
M	masculine
MID	medium
N	neuter
NA	not available
NEG	negative
NF	non-final
NMZ	nominalizer
NOM	nominative
NP	noun phrase
N-V	noun-verb
OBJ	object
OBL	oblique
OPT	optative
PASS	passive
PN	proper name
PP	prepositional phrase
PRET	preterit
PRF	perfect
PRS	present
PRT	particle
PRV	preverb
PST	past
PTCP	participle
POSS	possessive
REFL	reflexive
REL	relative
SEQ	sequence
SIM	simultaneous
SG	singular
STAT	stative
SUBJ	subjunctive
TAM	tense aspect mode
TMP	temporal
TOP	topic
QUOT	quotative
VOC	vocative
V-V	verb-verb

REFERENCES CITED

Alsina, A., Bresnan, J. & Sells, P. (Eds.). (1997). *Complex Predicates*. Stanford: CLSI.

Anklesaria, B. T. (1969). *The Pahlavi Rivāyat of Āturfarnbag and Farnbag-Srōš.* Bombay: K. L. Bhargva & Co. (=RAF)

Arkadiev, P. M. (2009). Poor (two-term) case systems: Limits on neutralization. In Malchukov, A. & Spencer, A. (Eds.). *Case*. Oxford: Oxford University Press.

Arunachalam, S. & Kothari, A. (2011). An experimental study of Hindi and English perfective interpretation. *Journal of South Asian Linguistics* 4.1, pp. 27-42.

Axenov, S. (2006). *The Balochi Language of Turkmenistan: A corpus-based grammatical description.* Uppsala: Acta Universitatis Upsaliensis.

Baker, M. (1988). *Incorporation: A theory of grammatical function changing.* Chicago: Chicago University Press.

Bakker, P., Daval-Markussen, A., Parkvall, M., & Plag, I. (2011). Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages* 26.1, pp 5-42.

Barðdal, J. & Eythórsson, T. (2012a). Reconstructing syntax: Construction grammar and the Comparative Method. In Boas, H. C., Sag, I. A. (Eds.). *Sign-based construction grammar*. (pp. 257-308). Stanford: CSLI Publications.

Barðdal, J. & Gildea, S. (2015). Diachronic Construction Grammar: Epistemological context, basic assumptions and historical implications. In Jóhanna Barðdal, Elena Smirnova, Lotte Sommerer, Spike Gildea. *Diachronic Construction Grammar*. (pp. 1-51). Amsterdam, Philadelphia: John Benjamins.

Baron I. & Herslund, M. (2001). Semantics of the verb HAVE. In: Irene Baron, Michael Herslund, & Finn Sørensen (eds.), *Dimensions of Possession*. Amsterdam, (pp. 85-99). Philadelphia: John Benjamins.

Barri, N. (1977). *Clause-models in Antiphontean Greek*. München: W Fink.

Benveniste, E. (1966). 'Etre' et 'avoir' dans leurs fonctions linguistiques. In: *Problèmes de linguistique générale*. (pp. 176-186). Paris: Gallimard. `

Borjian, H. (2014). What is Judeo-Median – and How Does It Differ from Judeo-Persian? *Journal of Jewish Languages* 2.2, pp. 117-142.

Bossong, G. (1985). *Differenzielle Objektmarkierung in den neuiranichen Sprachen.* Tübingen: Narr.

Bryant, D. & Moulton, V. (2004). Neighbor-Net: An Agglomerative Method for the Construction of Phylogenetic Networks. *Molecular Biology and Evolution* 21.2, pp. 255-265.

Boas, F. (1911). Introduction. In: *Handbook of American Indian languages*, vol. 1. (Bureau of American Ethnology, Bulletin 40.) Washington: Government Print Office.

Boas, H. C. (Ed.). (2010). *Contrastive studies in construction grammar*. Amsterdam, Philadelphia, John Benjamins.

Boas, H. C., & Sag, I. A. (Eds.). (2012). *Sign-based construction grammar*. Stanford: CLSI.

Bowerman, M. & Pederson, E. (1992). Topological relations picture series. In Levinson S. C (Ed.), Space stimuli kit 1.2: Nijmegen: Max Plank Institute for Psycholinguistics.

Buitenen, J. A (translator). (1973). *The Mahābhārata*. Chicago: Chicago University Press.

Butt, M. (1995). *The Structure of Complex Predicates in Urdu*. Stanford: CLSI.

Butt, M. (1997). Complex predicates in Urdu. In Alsina, A., Bresnan, J. & Sells, P (Eds.). *Complex Predicates*. (pp. 107-149) Stanford: CLSI.

Butt, M. & Lahiri, A. (2013). Diachronic pertinacity of light verbs. *Lingua 135*, pp. 7-29.

Bybee, J. (2001). *Phonology and language use*. Cambridge: Cambridge University Press.

Bybee, J. (2007). *Frequency of use and the organization of language*. Oxford, New York: Oxford University Press.

Cardona, G. (2003). Sanskrit. In Cardona, C. & Jain, D. (Eds.). *The Indo-Aryan Languages*. (pp. 104-160). London, New York: Routledge.

Carlier, A., De Mulder, W. & Lamiroy B. (2012). Introduction: The pace of grammaticalization in a typological perspective. *Folia Linguistica* 46.2, pp. 287-302.

Chappell, H. (2008). Variation in the grammaticalization of complementizers from *verba dicendi* in Sinitic languages. *Linguistic Typology* 12.1, pp. 45-98.

Cheung, J. (2007). *Etymological Dictionary of the Iranian Verb.* Boston MA: Brill.

Christian, I. (1987). *Language as a social behavior: Folk tales as a database for developing a beyond-the-discourse model.* Huntington Beach: Summer Institute of Linguistics.

Christmas, R. B. & Christmas J. E. (1973a). Kupia Texts. In R. L. Trail (Ed.) *Patterns In Clause, Sentence, and Discourse in selected languages of India and Nepal*. Katmandu: Tribhuvan University Press. (=Kupia)

Christmas, R. B. & Christmas, J. E. (1973b). Clause patterns in Kupia. In: Ronald L. Trail (ed.) *Patterns In Clause, Sentence, and Discourse in selected languages of India and Nepal II*. (pp. 257-345). Katmandu: Tribhuvan University Press.

Ciancaglini, C. A. (2011). The formation of the periphrastic verbs in Persian and neighboring languages. In Maggi, M. & Orsatti, P. (Eds.). *The Persian Language in history*. (pp. 3-21). Wiesbaden: Dr. Ludwig Reichert Verlag.

Cole, P., Harbert, W., Hermon, G. & Sridhar, S.N. (1980). The acquisition of subjecthood. *Language, 56*, pp. 719-743.

Comrie, B. (1978). Ergativity. In Lehmann, W. P. (Ed.). *Syntactic typology: Studies in the phenomenology of language*, (pp. 329–394). Austin: University of Texas Press.

Comrie, B. (1989). *Language universals and linguistic typology: Syntax and morphology*. 2nd edition. Oxford: Blackwell.

Clark, E. (1978). Locationals: Existential, locative, and possessive constructions. In Greenberg, J. (Ed.). *Universals of Human Language*. (pp. 85-126). Stanford: Stanford University Press.

Creissels, D. (2013). Control and the evolution of possessive and existential constructions. In Van Gelderen, E., Cennamo, M. & Barðdal, J. (Eds.). *Argument structure in flux*. (pp. 461-477). Amsterdam, Philadelphia: John Benjamins.

Creissels, D. (2014). Existential predication in typological perspective. Handout from *the 46th Annual Meeting of the Societas Linguistica Europaea*. Split, Croatia.

Croft, W. (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

Croft, W. (2009). Constructions and Generalizations. *Cognitive Linguistics* 20.1, 157-165.

Croft, W. (2013). Agreement as anaphora, anaphora as coreference. In Bakker, D. & Haspelmath, M. (Eds.). *Languages Across Boundaries: Studies in Memory of Anna Siewierska*. (pp. 107-129). Berlin: De Gruyter Mouton.

Croft, W. (2014). Comparing categories and constructions crosslinguistically (again): The diversity of ditransitives. *Linguistic Typology* 18.3, pp. 533-551.

Croft, W. (2016). Comparative Concepts and language-specific categories: theory and practice". *Linguistic Typology* 20.2, pp. 377-393.

Csardi, G. & Nepusz, T. (2006). The iGraph software package for complex network research. *Complex Systems* 1695, pp. 1-9.

Cysouw, M. (2014). Inducing semantic roles. In Luraghi, S. & Narrog, H. (Eds.). *Perspectives on semantic maps*. (pp. 23-68). Amsterdam, Philadelphia: John Benjamins.

Cysouw, M. & Good, J. (2013). Languid, Doculect and Glossonym: Formalizing the notion 'Language'. *Language Documentation & Conservation* 7, pp. 331-359.

Dain, D. (2003). Sociolinguistics of the Indo-Aryan languages. In Cardona, G. & Jain, D. (Eds.). *The Indo-Aryan Languages*. (pp. 46-66) London, New York: Routledge.

Delbrück, E. K. (1888). *Altindische Syntax.* Halle: Waisenhaus. Repr. 1968, Darmstadt: Wissenschaftlische Buchgesellschaft.

DeLancey, S. (2000). The Universal Basis of Case. *Logos and Language* 1.2, pp. 1-15.

DeLancey, S. (2011). Notes on verb agreement in Tibeto-Burman. *Himalayan Linguistics* 10.1, pp. 1-29.

Deo, A. & Sharma, S. (2006). Typological variation in the ergative morphology of Indo-Aryan languages. *Linguistic Typology* 10.3, pp. 369-418.

Dhakal, D. N. (2013). *Darai Texts*. München: Lincom.

Dhakal, D. N. (2015). Darai verb agreement. *Himalayan Linguistics* 14.2, 1-38.

Dryer, M. S. & Haspelmath, M. (2013). *The World Atlas of Language Structures*. Leipzig: Max Plank Institute for Evolutionary Anthropology.

Du Bois, J. (1987). The discourse basis of ergativity. *Language* 63, pp. 805-855.

Du Bois, J., Kumpf, L., & Ashby, W. (2003). *Preferred Argument Structure.* Amsterdam, Philadelphia: John Benjamins.

Durkin-Meisterernst, D. (2014). *Grammatik des Westmitteliranischen (Partisch und Mittelpersisch)*, Vienna: Verlag der Österreichischen Akademie der Wissenschaften.

Enfield, N. J. (2003). *Linguistic epidemiology: Semantics and grammar of language contact in mainland Southeast Asia.* Abingdon, New York: RoutlageCurzon.

Fillmore, C. J. (1967). The case for case. In Bach, E. & Harms, R. T. (Eds.). *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston.

Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph Drawing by Force-Directed Placement. Software: Practice and Experience, 21(11), 1129-1164.

Garbutt, K. (2006). *Mahābhārata: book four*. New York: New York University Press.

Ghomeshi, J. & Massam, D. (1994). Lexical/syntactic relations without projection. *Linguistic Analysis* 24(3-4), pp. 175-217.

Gindin, T. (2007). *The Early Judaeo-Persian tafsīrs of Ezekiel: text, translation, commentary* (3 vols). Vienna: Verlag der Österrichischen Akademie der Wissenschaften.

Gildea, S. (1998). *On reconstructing grammar: comparative Cariban morphosyntax.* Oxford: Oxford University Press.

Givón, T. (1979). *On understanding grammar*. New York: Academic Press.

Givón, T. (2001). *Syntax: An introduction. 2 vols.* Amsterdam: Benjamins.

Givón, T. (2008). *The genesis of syntactic complexity.* Amsterdam, Philadelphia: John Benjamins.

Givón, T. (2013). On the diachrony of the 'Ethical Dative'. In Thornes, T., Andvik, E., Hyslop, G., & Jansen, J. (Eds.). *Functional-Historical Approaches to Explanation: In honor of Scott DeLancey*. (pp. 43-66). Philadelphia, Amsterdam: John Benjamins.

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: University of Chicago Press.

Goldberg, A. (2003). Words by default: The Persian Complex Predicate Construction. In Francis, E. J. & Michaelis, L. A. (Eds.). *Mismatch: Form-function incongruity and the architecture of grammar.* (pp. 117-146). Stanford: CLSI.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language.* Oxford: Oxford University Press.

Greenberg, J. H. (1963). Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, J. H. (Ed.). *Universals of Language.* (pp. 73-113). Cambridge, Mass.: MIT Press.

Gustafsson, U. (1973a). Kotia Oriya Texts. In R. L. Trail (ed.) *Patterns In Clause, Sentence, and Discourse in selected languages of India and Nepal*. Katmandu: Tribhuvan University Press.

Gustafsson, U. (1973b). Clause patterns in Kotia Oriya. In Trail R. L (Ed.). *Patterns in Clause, Sentence, and Discourse in selected languages of India and Nepal II*. (pp. 191-257). Katmandu: Tribhuvan University Press.

Haig, G. (2002). Noun-plus-verb complex predicates in Kurmanjî Kurdish: Argument sharing, argument incorporation, or what? *STUF-Language Typology and Universals*. 55.1, pp. 15-48.

Haig, G. (2008). *Alignment change in Iranian languages: A construction grammar approach.* Berlin: De Gruyter Mouton.

Haig, G. & Öpengin, E. (2014). Introduction to Special Issue- Kurdish: A critical research overview. *Kurdish Studies* 2.2, pp. 99-122.

Haig, G. & Schnell, S. (2016). The discourse basis of ergativity revisited. *Language* 92.3, pp: 591-618.

Haig, G. & Theile, H. (2014). Post-predicate Goals in Northern Kurdish and neighboring languages: a pilot study in quantitative areal linguistics. Handout. Presentation at the *Second International Conference on Variation and Change in Kurdish*.

Hammarström, H., Forkel, R., & Haspelmath, M. (2017). *Glottlog 3.0* Jena: Max Plank Institute for the Science of Human History.

Harris A. C. & Campbell, L. (1995). *Historical syntax in cross-linguistic perspective*. Cambridge: Cambridge University Press.

Hartmann I., Haspelmath, M. & Cysouw, M. (2014). Identifying semantic role clusters and alignment types via microrole coexpression tendencies. *Studies in Language* 38, pp. 463-484.

Haspelmath, M. (2004). Directionality in language change, with special reference to unidirectionality in grammaticalization. In Fischer, O., Norde, M. & Perridon, H. (Eds.). *Up and down the cline*. (pp. 17-44). Amsterdam, Philadelphia: John Benjamins.

Haspelmath, M. (2005). Argument marking in ditransitive alignment types *Linguistic Discovery 3.1,* pp. 1-21.

Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. Language 86.3, pp. 663-687.

Haspelmath, M. (2011). On S, A, P, T, and R as comparative concepts for alignment typology. *Linguistic Typology* 15.3, pp. 535-567.

Haspelmath, M. (2013). Argument indexing: a conceptual framework for the syntactic status of bound person forms. In Bakker, D. & Haspelmath, M. (Eds.). *Languages Across Boundaries: Studies in Memory of Anna Siewierska*. (pp. 197-226). Berlin: De Gruyter Mouton.

Haspelmath, M. (2015). Transitivity Prominence. In Malchukov, A. & Comrie, B. (Eds.). *Valency Classes in the World's Languages: Introducing the framework, and case studies from Africa and Eurasia.* (pp. 131-149). Berlin, Boston: De Gruyter Mouton.

Haspelmath, M. (2016). The Serial Verb Construction: Comparative Concept and Cross-Linguistic Generalizations. *Language & Linguistics* 17.3, pp. 291-319.

Heine, B. (1992). Grammaticalization Chains. *Studies in Language* 16, pp: 335-368.

Heine, B. (1997). *Possession: Cognitive sources, forces, and grammaticalization*. Cambridge: Cambridge University Press.

Heine, B. (1993). *Auxiliaries*. Oxford: Oxford University Press.

Heine, B. (2000). Grammaticalization chains across languages: An example from Khoisan. In Gildea, S. (Ed.). *Reconstructing Grammar: Comparative Linguistics and Grammaticalization*. (pp. 177-200). Amsterdam, Philadelphia: John Benjamins.

Heine, B. & Kuteva, T. (2002). *World lexicon of grammaticalization.* Cambridge: Cambridge University Press.

Herring, S. C. (1988). Aspect as a Discourse Category in Tamil. *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistic Society*, pp. 280-292.

Heston, W. L. 1976. *Selected Problems in fifth to tenth century Iranian syntax.* University of Pennsylvania: Unpublished Ph.D Thesis.

Hilpert, M. 2010. What can synchronic gradience tell us about reanalysis? Verb-first conditionals in written German and Swedish. In Traugott, E. C. & Trousdale, G.

(Eds.). *Gradience, Gradualness, and Grammaticalization*. (pp. 181-201). Amsterdam, Philadelphia: John Benjamins.

Hockett, C. F. (1955). Attribution and apposition" *American Speech* 30.2, pp. 99-102.

Hook, E. P. (1975). *Compound Verbs in Hindi*. PhD thesis, University of Michigan.

Hook, E. P. (1991). The Emergence of Perfective Aspect in Indo-Aryan Languages. In: Traugott, E. C. & Heine, B. (Eds.). *Approaches to Grammaticalization Volume 2.* (pp. 59-89). Philadelphia, Amsterdam: John Benjamins.

Hook, E. P. (1995). The Compound Verb in Gujarati and its Use in Connected Text. In: D. T. Vyas (ed.), *Studies in Jaina Art and Iconography and Allied Subjects in Honour of Dr. U.P. Shah.* (pp. 339-356). Vadodara: Abhinav Publications.

Hopper, P. J. & Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language* 56.2, pp 251-299.

Hopper, P. J. & Traugott, E. C. (1993). *Grammaticalization*. 2nd Edition. Cambridge: Cambridge University Press.

Hudson, D. H. & Bryant, D. (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* 23, pp 254-267.

Jackendoff, R. (1983). *Semantics and Cognition*. Cambridge, MA: MIT Press.

Jamaspasa, K. M. & Humbach, H. (1971). *Pursišnīhā: A Zoroastrian catechism.* Weisbaden: Reichert.

Jeremiás, É. M. (1984). Diglosia in Persian. *Acta Linguistica Academiae Scientiarum Hungaricae*. 34.

Johanson, L. (1995). On Turkic converb clauses. In Haspelmath, M. & König, E. (Eds.). *Converbs in cross-linguistic perspective: Structure and meaning of adverbial verb forms – adverbial participles, gerunds*. (pp. 313-348). Berlin: Mouton de Gruyter.

Joseph, B. D. (2014). What counts as (an instance of) grammaticalization? *Folia Linguistica* 48.2, pp. 361-384.

Joseph, B. D. & Pappas, P. (2002). On some recent views concerning the development of the Greek future system. *Byzantine and Modern Greek Studies* 26.1, pp. 247-273.

Josephson, J. (2016). The expression of modality in Late Middle Persian. In: Jila Ghomeshi, Carina Jahani, Agnès Lenepveu-Hotz (eds.) *Further Topics in Iranian Linguistics*. (pp. 71-90). Paris: L'association pour l'avancement des études Iraniennes.

Kachru, Y. (2006). *Hindi*. Amsterdam, Philadelphia: John Benjamins.

Korn, A. (2009). Western Iranian Pronominal Clitics. *Orientalia Suecana LVIII*, pp. 159-171.

Korn, A. (2013). Looking for the middle way: Voice and transitivity in complex predicates in Iranian. *Lingua* 135, pp. 30-55.

Korn, A. 2015. Review of: Nicholas Sims-Williams 2012: Bactrian Documents from Northern Afghanistan I: Legal and Economic Documents, revised edition. Bactrian Documents from Northern Afghanistan III: Plates. *Indo-Iranian Journal* 58/1, pp. 87-91

Korn, A. (2016). A partial tree of Central Iranian: A new look at Iranian subphyla. *Indogermanische Forschungen* 121.1, pp. 401-434.

Krishnamurti, B. (1986). *South Asian Languages: Structure, Convergence, and Diglosia.* Delhi, Varanasi, Patna, Madras: Motilal Banarasidass.

Lambrecht, K. (1988). There was a farmer had a dog: Syntactic Amalgams Revisited. *Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistic Society*, pp. 319-339.

Lamiroy, B. (1999). Auxiliaires, langues romanes et grammaticalisation. *Langages* 135, pp. 33-45.

Lamiroy, B. (2003). Grammaticalisation et comparasion de langues. *Verbum* 25, pp. 411-431.

Langacker, R. W. (1977). Syntactic reanalysis. In Li, N. L. (Ed.). *Mechanisms of syntactic change*. (pp. 57-141). Austin: University of Texas Press.

Langacker, R. W. (1993). Reference-point constructions. *Cognitive Linguistics* 4.1, pp. 1-38.

Lazard, G. (1963). *La langue des plus anciens monuments de la prose persane*. Paris: Klincksieck.

Lazard, G. (1994). *L'actantce.* Paris: Presses Universitaires de France.

Lazard, G. (2002). Transitivity revisited as an example of a more strict approach in typological research. *Folia Linguistica* 36(3-4), pp. 141-190.

Lazard, G. (2015). Two possible universals: The Major Biactantial Construction; the twofold notion of subject. *Linguistic Typology* 19.1. pp. 111-130.

Lecoq, P. (1979). *Le Dialecte de Sivand*. Weisbaden: Dr. Ludwig Reichert Verlag.

Levin, B. (1993). *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago: University of Chicago Press.

Lewis, P. M., Simmons, G. F. & Fennig, C. D. (2009). *Ethnologue: Languages of the World. Dallas*. Texas: SIL International. Online version.

Liljegren, H. (2010). Where have all the verbs gone? On verb stretching and semi-words in Indo-Aryan Palula. *Himalayan Linguistics* 9.1, pp. 51-79.

Liljegren, H. (2016). *A Grammar of Palula.* Berlin: Language Science Press.

Liljegren, H. & Haider, N. (2015). *Palula Texts*. Islamabad, Pakistan: Forum for Language Initiatives.

Mahmoudveysi, P., Bailey, D., Paul, L., & Haig. G. (2012). *The Gorani language of Gawrajū, a village of West Iran.* Weisbaden: Dr. Ludwig Reichert Verlag.

Macdonell, A. (1958 [1916]). *A Vedic grammar for students*. Oxford: University Press.

MacKenzie, D. N. (1961). *Kurdish dialect studies*. *Vol I*. London: Oxford University Press.

MacKenzie, D. N. (1976). *The Buddhist Sogdian texts of the British Library*. Téhéran, Liege: Brill.

Malchukov, A. L., Haspelmath, M. & Comrie, B. (2010). Ditransitive construction: A typological overview. In Malchukov A. L., Haspelmath, M. & Comrie, B. (Eds.). *Studies in ditransitive constructions: A comparative handbook*. (pp. 1-64). Berlin: De Gruyter Mouton.

Masica, C. P. (1976). *Defining a Linguistic Area: South Asia.* Chicago: Chicago University Press.

Masica, C. P. (1991). *The Indo-Aryan Languages*. Cambridge: Cambridge University Press.

Megerdoomian, K. (2012). The status of the nominal in Persian complex predicates. *Natural Language and Linguistic Theory* 30.1, pp. 179-216.

Michaelis, S. M., Maurer, P., Haspelmath, M. & Huber, M. (2013). *Atlas of Pidgin and Creole Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.

Miner, K. L. (1986). Noun stripping and loose incorporation in Zuni. *International Journal of American Linguistics* 52.3, pp. 242-252.

Mithun, M. (1984). The evolution of noun incorporation. *Language* 60.4, pp. 847-894.

Mohanan, T. (1994). *Argument Structure in Hindi*. Stanford: CLSI.

Mohanan, T. (1997). Multidimensionality of representation: NV complex predicates in Hindi. In Alsina, A., Bresnan, J. & Sells, P. (Eds.). *Complex Predicates*. (pp. 431-471). Stanford: CLSI.

Mohanan, K. P. & Mohanan, T. (1994). Issues in Word Order in South Asian Languages: Enriched Phrase Structure of Multidimensionality?, In Butt, M., Holloway, T. K. & Ramchand, G. (Eds.). *Theoretical Perspectives on Word Order in South Asian Languages*. (pp. 153-184). Stanford: CLSI.

Montaut, A. (2012). *Le Hindi.* Leuven, Paris: Peeters.

Morgenstierne, Georg. (1974) *Etymological dictionary of the Shughni group.* Wiesbaden: Reichert.

Müller, S. (2010). Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46(03), pp. 601-655.

Müller, S. (2016). *Grammatical Theory: From Transformational Grammar to Constraint-Based Approaches*. Language Science Press.

Næss, Å. (2007). *Prototypical Transitivity*. Amsterdam, Philadelphia: John Benjamins.

Neukom, L. & Patnaik, M. (2003). *A grammar of Oriya*. Zürich: Seminar für Allgemeine Sprachwissenschaft der Univ. Zürich.

Newmeyer, F. J. (2007). Linguistic typology requires crosslinguistic formal categories. *Linguistic Typology* 11(1), pp. 133-157.

Oberlies, T. (2003). Aśokan Prakrit and Pāli. In Cardona, G. & Jain, D. (Eds.). *The Indo-Aryan Languages*. (pp. 161-203). London, New York: Routledge.

Paul, D. (2011). *A Comparative Dialectal Description of Iranian Taleshi*. A thesis submitted to the university of Manchester for the degree of PhD.

Paul, L. (2007). Zur Lage der Gōranī-dialekte im Iran und ihrer Erforschung. In Macuch, M., Maggi, M. & Sundermann, W. (Eds.), *Iranian languages and Texts from Iran and Turan*. *Ronald E. Emmerick Memorial Volume*. Wiesbaden: Harrassowitz.

Paul, L. (2013). *A grammar of early Judaeo-Persian*. Wiesbaden: Reichert Verlag.

Payne, J. R. (1980). The decay of ergativity in Pamir languages. *Lingua* 51(2-3), pp. 147-186.

Payne, T. E. (1997). *Describing morphosyntax: a guide for field linguists*. Cambridge: Cambridge University Press.

Payne, D. L. (2009). Is possession mere location? Contrary evidence from Maa. In McGregor, W. B. (Ed.). *The Expression of Cognitive Categories (ECC): The Expression of Possession*. (pp. 107-143). Berlin: De Gruyter Mouton.

Pederson, E. (2008). Event realization in Tamil. In: Bowerman, M. (Ed.). *Crosslinguistic Perspectives on Argument Structure.* (pp. 331-355). New York, London: Lawrence Erlbaum Associates.

Peterson, J. M. (1998). *Grammatical Relations in Pāli and the Emergence of Ergativity in Indo-Aryan*. München: Lincom.

Plank, F. (1995). (Re-)Introducing Suffixaufnahme. In Plank, F. (Ed.). *Double Case*: *Agreement by Suffixaufnahme*. (pp. 3-113)New York, Oxford: Oxford University Press.

Pokorny, J. (1959). *Indogermanisches etymologisches Wörterbuch Vol I-II.* Bern: Francke.

Poornima, S. (2012). Hindi Complex Predicates at the Syntax-Semantics Interface. PhD dissertation, University of Buffalo.

Premchand, M. (2017[1936]). *Godaan* (Gift of a cow). San Bernadino: Durlabh eSahitya corner.

R Development Core Team. (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, retrieved 16 May, 2011 from http://www.R-project.org/

Samvelian, P. (2012). *Grammaire des prédicats complexes. Les constructions nom-verbe.* Paris: Hermès-Lavoisier.

Samvelian, P. & Faghiri, P. (2013). Re-thinking Compositionality in Persian Complex Predicates. *Proceedings of the 39th annual meeting of the Berkeley Linguistic Society,* pp. 212-226.

Schliep, K.P. 2011. Phangorn: Phylogenetic analysis in R. *Bioinformatics* 27.4, pp. 592-593.

Shaked, S. (1979). *The Wisdom of the Sasanian Sages (Dēnkard VI)*. Boulder Co.: Westview Press.

Shapiro, M. C. (2003). *Hindi*. In Cardona, G. & Jain, D. (Eds.). *The Indo-Aryan Languages*. (pp. 250-285). London, New York: Routledge.

Shibatani, M. (2009). On the form of complex predicates: towards demystifying serial verbs. In Helmbrecht, J., Nishina, Y., Shin, Y. M., Skopeteas, S. & Verhoeven, E. (Eds.). *Form and Function in Language Research: Papers in Honour of Christian Lehmann.* (pp. 255-282). Berlin, New York: De Gruyter Mouton.

Shirtz, S. (2016). Indirect participants as core-arguments in Middle Persian. In: Jila Ghomeshi, Carina Jahani, Agnès Lenepveu-Hotz (eds.) *Further Topics in Iranian Linguistics*. (pp. 175-195). Paris: L'association pour l'avancement des études Iraniennes.

Shokri, G., Jahani, C. & Barani, H. (2013). *When Tradition Meets Modernity: Five Life Stories from the Galesh Community in Ziarat, Golestan, Iran.* Uppsala: Acta Universitatis Upsaliensis.

Sims-Williams, N. (1981). Notes on Manichean Middle Persian Morphology. *Studia Iranica* 10, pp. 165-176.

Sims-Williams, N. (1982). The double system of nominal inflexion in Sogdian. *Transactions of the Philological Society* 80.1, pp. 67-76.

Sims-Williams, N. (1984). The Sogdian 'Rhythmic Law'. Skalmowski, W & Van Tongerloo, A. (Eds.). *Middle Iranian Studies: Proceedings of the International Symposium Organized by the Katholieke Univesiteit Leuven*. (pp. 203-215). Leuven: Peeters.

Sims-Williams, N. (1990). Chotano-Sogdica II: Aspects of the Development of nominal morphology in Khotanese and Sogdian. In: *Proceedings of the first European Conference of Iranian Studies*. (pp. 275-296). Rome.

Sims-Williams, N. (2000). *Bactrian Documents from Northern Afghanistan, I: Legal and Economic Documents*. Oxford: Oxford University Press.

Sims-Williams, N. (2007). *Bactrian Documents from Northern Afghanistan, II: Letters and Buddhist texts*. Oxford: Oxford University Press.

Sims-Williams, N. (2008). The Sogdian Potentialis. In Macuch, M., Maggi, M. & Sundermann, W. (Eds.). *Iranian languages and texts from Iran and Turan. Ronald E. Emmerick memorial volume*. (pp. 377-386). Wiesbaden: Harrassowitz.

Sims-Williams, N. (2010). Differential Object Marking in Bactrian. In: Korn, A., Haig, G., Karimi, S. & Samvelian, P. (Eds.). *Topics in Iranian Linguistics*. (pp. 23-38). Wiesbaden: Reichert.

Skjærvø, O. P. (2009). Middle West Iranian. In: Windfuhr, G. (Ed.). *The Iranian Languages*. (pp. 196-278). London, New York: Routledge.

Slade, B. (2013). The diachrony of light and auxiliary verbs in Indo-Aryan. *Diachronica 30.4,* pp 531-578.

Sreedhar, M.V. (1985). *Standardized Grammar of Naga Pidgin.* Mysore: CIIL.

Staal, J.F. (1967). *Word Order in Sanskrit and Universal Grammar*. Dordrecht: D. Reidel.

Stassen, L. (1985). *Comparison and Universal Grammar*. Oxford: Blackwell.

Stassen, L. (1997). *Intransitive Predication*. Oxford: Oxford University Press.

Stassen, L. (2009). *Predicative Possession*. Oxford: Oxford University Press.

Stassen, L. (2013a). Nominal and Locational Predication. In Dryer, M. S. & Haspelmath, M. (Eds.). *The World Atlas of Language Structures*. Leipzig: Max Plank Institute for Evolutionary Anthropology.

Stassen, L. (2013b). Predicative Possession. In Dryer, M. S. & Haspelmath, M. (Eds.). *The World Atlas of Language Structures*. Leipzig: Max Plank Institute for Evolutionary Anthropology.

Steels, L. (2011). Introducing Fluid Construction Grammar, in Steels, L. (Ed.). *Design Patterns in Fluid Construction Grammar*. (pp. 3-31). Amsterdam, Philadelphia: John Benjamins.

Stilo, D. (2004). Double-Duty Pronominal Clitics in Gazi. Handout from *Syntax of the World's Languages.*

Stilo, D. (2007). Isfahan, Jewish Dialects. *Encyclopædia Iranica*. (online publication).

Stilo, D. (2009). Case in Iranian: From reduction and loss to innovation and renewal. In: Malchukov, A. & Spencer, A (Eds.). *Case*. (pp. 700-715). Oxford: Oxford University Press.

Tapas, R. S. (2003). Oriya. In: George Cardona & Dhanesh Jain (eds.) *The Indo-Aryan Languages*. London, (pp. 444-476). New York: Routledge.

Tedesco, P. (1921). Dialektologie der mitteliranischen Turfantexte. *Monde Oriental* 15, pp. 184-258.

Timberlake, A. (1977). Reanalysis and actualization in syntactic change. In Li, C. N. (Ed.). *Mechanisms of syntactic change*. (pp. 141-180). Austin: University of Texas Press.

Traugott, E. C. & Trousdale, G. (2010). *Gradience, Gradualness, and Grammaticalization*. Amsterdam, Philadelphia: John Benjamins.

Trudgill P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.

Upadhaya, S. K. (tr.). (2011). *Nagamese Baibel*. Kohima: Nagamese Bible Publication

Vahman, F. (1988). *Ardā Wirāz Nāmag- The Iranian 'Divina Commedia',* London, Malmo: Curzon Press.

Van de Velde, F. & Lamiroy, B. (2016). External possessors in West Germanic and Romance: Differential speed in the drift towards NP configurationallity. In Cuyckens, H., Ghesquière, L. & Van Olman, D. (Eds.). *Aspects of Grammaticalisation: (Inter)Subjectivity, Analogy, and Unidirectionality.* (pp. 353-399). Berlin: Mouton de Gruyter.

Vendler, Z. (1967). Events and time. *Linguistics and Philosophy* 37.

Verbeke, S. (2013). *Alignment and Ergativity in New Indo-Aryan Languages*. Berlin: De Gruyter Mouton.

Verkerk, A. (2014). *The evolutionary dynamics of motion event encoding.* Ph.D dissertation, Radboud Universiteit Nijmegen.

Wälchli, B. & Cysouw, M. (2012). Lexical typology through similarity semantics: Toward a semantic map of motion verbs. *Linguistics* 50.3, pp. 671-710.

Warns, G. R. et al. (2015). *Gplots: Various R Programming Tools for Plotting Data*.

Wasserman, S. & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. New York: Springer Verlag.

Williams, A. V. (1990). *The Pahlavi Rivāyat Accompanying the Dādestān ī dēnīg*. Copenhagen: Munksgaard.

Yakubovich, I. S. (2010). Indo-European **mā* 'to grow'. *Indo-European Linguistics and Classical Philology*. 14.2, pp 478-492

Yoshida, Y. 2009. Sogdian. In Windfuhr, G. (Ed.). *The Iranian Languages.* (pp. 279-335). London, New York: Routledge.