

**Conceptual and Procedural Encoding in Relevance Theory:  
A Study with Reference to English and Kiswahili**

Stephen M. Nicolle

Thesis submitted for the Degree of  
Doctor of Philosophy

University of York

Department of Language and Linguistic Science

August 1996

## Abstract

Recent work within relevance theory suggests that the distinction between truth-conditional and non-truth-conditional meaning is of limited value in the study of linguistic (as opposed to logical) semantics. Of potentially greater interest is the distinction between conceptual and procedural encoding. Conceptual encoding contributes to the construction of conceptual representations in the mind of the addressee, whilst procedural encoding provides the addressee with instructions as to how conceptual representations are to be manipulated to achieve relevance.

The primary aims of this thesis are, firstly to clarify the currently intuitive distinction between conceptual and procedural encoding by providing a precise characterisation of procedural information and a set of criteria by which linguistically encoded information can be identified as either conceptual or procedural, and secondly to determine some of the implications of the distinction for linguistic semantics.

I propose that procedural encoding is of two basic kinds: that which constrains the manipulation of propositional conceptual representations, and that which constrains the manipulation of sub-propositional conceptual representations. An example of the former is the discourse connective *so*, which establishes an inferential connection between the proposition with which it is associated and some highly accessible assumption. This type of procedural encoding has been widely discussed (see for example Blakemore 1987, 1988).

Less widely discussed is procedural encoding which constrains the manipulation of sub-propositional conceptual representations. I propose that such encoding is a property of grammaticized expressions, such as pronouns, and tense, aspect and modality markers. I demonstrate how a procedural account of grammatical markers accounts not only for their role in utterance interpretation (constraining the construction of propositional conceptual representations) but also for their historical development through grammaticization, and the corresponding variation in usage which linguistic change engenders.

I illustrate my thesis through analysis of constructions in English (notably *will*, *shall* and *be going to*) and Kiswahili.



1391680

<b>CONTENTS</b>	<b>page</b>
<b>Abstract</b>	2
<b>Acknowledgement</b>	6
<b>Introduction</b>	7
0.1 Aims of the thesis	7
0.2 Why relevance theory?	9
0.3 Overview	12
<b>Chapter 1 Relevance Theory: An Introduction</b>	<b>13</b>
<b>1.1 Inferential Communication</b>	<b>13</b>
1.1.1 The primacy of inferential over coded communication	14
1.1.2 Three phases of utterance interpretation	16
<b>1.2 Phase One: Linguistic Decoding</b>	<b>17</b>
1.2.1 The modularity hypothesis	18
1.2.2 The role of linguistic decoding in the identification of propositional form	19
1.2.3 Assessing the weakly interactive model	23
1.2.4 Summary	26
<b>1.3 Phase Two: Inferential Enrichment</b>	<b>27</b>
1.3.1 Logical form	27
1.3.2 Disambiguation, reference assignment and inferential enrichment	29
1.3.3 Gricean pragmatics and relevance theory	31
<b>1.4 Phase Three: Relevance</b>	<b>35</b>
1.4.1 Contextual effects and processing effort	35
1.4.2 The principle of relevance	36
1.4.3 Consistency with the principle of relevance	39
1.4.4 Summary	42
<b>Chapter 2 Conceptual and Procedural Encoding</b>	<b>44</b>
<b>2.1 Information Types in Linguistic Semantics</b>	<b>44</b>
2.1.1 Arguments for a dualistic semantics	45
2.1.2 A reappraisal of the conceptual/procedural distinction	50
<b>2.2 Exemplification of Procedural Encoding at Phase Three</b>	<b>55</b>
2.2.1 Constraints on implicatures 1: Discourse connectives <i>so</i> and <i>well</i>	56
2.2.2 Constraints on implicatures 2: Information packaging	67
2.2.3 Constraints on higher-level explicatures: Kiswahili <i>je/-je</i> interrogatives	75

2.3	<b>Conclusions</b>	81
2.3.1	Procedural encoding at phase three: a theoretical characterisation	81
2.3.2	Procedural encoding at phase three: a descriptive characterisation	83
2.3.3	Summary	85
<b>Chapter 3</b>	<b>Procedural Encoding at Phase Two</b>	86
3.1	<b>Pronouns</b>	87
3.1.1	A procedural account of pronouns	88
3.1.2	Accessibility	94
3.1.3	The Kiswahili Object Marker as an exponent of procedural encoding	97
3.1.4	Conclusion	110
3.2	<b>Situations and Events</b>	111
3.2.1	Situation representations and operators	112
3.2.2	Domains	116
3.2.3	Summary	118
3.3	<b>Tense, Aspect and Time</b>	119
3.3.1	Temporal reference	119
3.3.2	Tense	122
3.3.3	Aspect and future time reference	126
3.4	<b>Summary and Prospects</b>	131
3.4.1	Summary: Pronouns, Tense and Aspect	131
3.4.2	Prospects: Modality	133
<b>Chapter 4</b>	<b>The English Modal Auxiliaries</b>	134
4.1	<b>Theoretical Preliminaries</b>	134
4.1.1	Logical analyses of the modals	134
4.1.2	Traditional descriptive analyses of the modals	140
4.1.3	Summary	141
4.2	<b><i>Can, May, Must and Should</i></b>	142
4.2.1	Modals as procedural markers of potentiality: (Klinge 1993)	143
4.2.2	Modals as exponents of conceptual encoding: (Papafragou 1995)	146
4.2.3	Towards a procedural analysis of the modals: (Groefsema 1995a)	150
4.2.4	A procedural analysis of <i>can, may must and should</i>	153
4.3	<b>Summary and Prospects</b>	156
4.3.1	Summary	156
4.3.2	Prospects	157

<b>Chapter 5</b>	<b>Grammaticization and Procedural Encoding</b>	<b>159</b>
5.1	<b>The gradualness of grammaticization</b>	160
5.1.1	The consequences of semantically gradual grammaticization	161
5.1.2	Counter-evidence from verb-serialisation	163
5.1.3	Counter-evidence from Kiswahili	165
5.1.4	Summary	169
5.2	<b>The causes of grammaticization</b>	170
5.2.1	Metaphorical extension and inference	171
5.2.2	Generalisation, harmony and absorption	172
5.2.3	The mechanism of semantic change in grammaticization	173
5.3	<b>Semantic retention</b>	176
5.3.1	Problems for traditional analyses	177
5.3.2	Semantic retention: the case of <i>will</i>	178
5.4	<b>Summary</b>	185
<b>Chapter 6</b>	<b>Future Expressions in English: <i>will</i>, <i>shall</i> and <i>be going to</i></b>	<b>187</b>
6.1	<b><i>Will</i></b>	187
6.1.1	<i>Will</i> and modality	188
6.1.2	<i>Will</i> and temporal reference	189
6.2	<b><i>Will</i> and <i>shall</i></b>	192
6.2.1	Agent-events and world-events	192
6.2.2	<i>Shall</i>	195
6.3	<b><i>Will</i> and <i>be going to</i>: syntactic differences</b>	197
6.3.1	<i>Be going to</i> as an operator	199
6.3.2	Co-occurrence restrictions with <i>already</i>	204
6.4	<b><i>Will</i> and <i>be going to</i>: semantic differences</b>	207
6.4.1	Procedural information encoded by <i>be going to</i> : a justification	207
6.4.2	A descriptive characterisation of <i>be going to</i>	209
6.4.3	The pragmatics of <i>be going to</i>	213
6.4.4	The conceptual semantic content of <i>be going to</i>	217
6.5	<b>Summary</b>	219
<b>Chapter 7</b>	<b>General Conclusion</b>	<b>223</b>
7.1	The theoretical basis	223
7.2	Procedural encoding at phase two	224
7.3	A diachronic perspective	225
7.4	Prospects	226
	<b>Key to Kiswahili sources</b>	<b>227</b>
	<b>Bibliography</b>	<b>227</b>

## Acknowledgement

This thesis could not have been completed without the support and encouragement of many people; I should especially like to thank the following:

I am grateful to my supervisor during the last two years, Dr. David Adger, whose detailed and critical comments facilitated vast improvements to the thesis. His confidence in me was a source of strength and encouragement when my own confidence failed.

I am also grateful to Dr. Joan Russell who supervised the first year of the thesis, not least for stimulating my interest in Kiswahili. Her comments during my upgrading, together with those of Steve Harlow, were greatly appreciated.

I owe special debt of thanks to all the staff of the Department of Language and Linguistic Science at the University of York, where I have been privileged to work in a friendly atmosphere of intellectual stimulation and open debate. I am particularly indebted to the Head of Department, Dr. Anthony Warner, for inviting me back to teach, and for the financial support afforded over the last two years by a Departmental Graduate Scholarship.

I have also benefited from discussion with researchers from elsewhere, some of whom are mentioned in the thesis. I am especially grateful to Dr. Marjolein Groefsema who provided stimulating feedback on an early draft of the thesis.

Finally, the biggest debt of thanks is owed to my family: to my parents, Anthony and Mary, for their unfailing support - both moral and financial, and to my wife Alison. Without her patient encouragement I doubt that the thesis would have been completed at all, and so it is dedicated to her.

## 0 INTRODUCTION

### 0.1 Aims of the thesis

This thesis explores the relevance theoretic distinction between conceptual and procedural encoding, with the aim of explicating and testing this potentially important distinction. The central distinction underlying theories of meaning until recently has been the distinction between truth-conditional and non-truth-conditional meaning; however this has failed to account for a range of linguistic phenomena (some of which are discussed in this thesis) and new approaches are being explored. It has been suggested (Sperber & Wilson 1995:259) that the distinction between conceptual and procedural encoding may prove to be of greater importance for linguistic semantics than the traditional distinction between truth-conditional and non-truth-conditional meaning.

This suggestion reflects the widely accepted view that cognitive science provides a better framework for the study of utterance interpretation than logic does. Whilst the distinction between truth-conditional and non-truth-conditional meaning is a logical distinction, the conceptual/procedural distinction is based in general principles of cognitive science. All current cognitive theories recognise the following two facts: first, that information processing involves a certain amount of effort on the part of the processor; and secondly, that an increase in computational effort reduces the chances of an information processing task being completed successfully. Now, human beings are rational information processors; rationality, in this context, involves not only the ability to derive valid conclusions from premises but also “the ability to allocate one’s cognitive resources efficiently.” (Sperber *et al* 1995:44) As a result, most cognitive theories (including Sperber & Wilson’s (1986, 1995) relevance theory, Johnson-Laird’s (1983) mental models theory, and Fodor’s (1983) modularity hypothesis) assume that information processing by humans is driven, on the one hand by the need to achieve successful outcomes, and on the other by the need to do so as efficiently as possible.<sup>1</sup>

Utterance interpretation is a form of information processing in which the information to be processed is both ostensively communicated (that is, it is intentional and conveys an expectation that it will be worth processing) and linguistically encoded. Given that utterances are produced and interpreted by rational information processors, and that it is in the interests of both speakers and addressees that utterance interpretation should be successful, we would expect utterances to be structured in

---

<sup>1</sup> Principles of least effort are invoked in other cognitively influenced theories, such as the Minimalist Programme in syntax (Chomsky 1991, 1993, 1995).

such a way as to reduce to a minimum the processing effort required. This is the basis of the conceptual/procedural distinction in relevance theory.

Blakemore (1987, 1988a) posited a distinction between two types of linguistically encoded information: conceptual, which contributes to propositions, viewed in relevance theory as conceptual representations; and procedural, which constrains the inferential processing of propositions, viewed as psychological computations performed over conceptual representations. In constraining the inferential processing of propositions, procedural information reduces the computational effort required of the addressee, thereby aiding the utterance interpretation process. As exponents of procedural encoding, Blakemore discussed non-truth-conditional discourse connectives such as *so* and *after all*.

It is also recognised in relevance theory that inference is involved not only in the processing of propositions during utterance interpretation, but also in the identification of those propositions. Wilson & Sperber (1981) demonstrated that propositions are not identified simply through linguistic decoding, but that inferential processing is crucial to disambiguation, resolution of vagueness, reference assignment, and the determination of ellipsed material, all of which help determine truth-conditions, and hence contribute to the proposition expressed. Wilson & Sperber (1993a) then suggested that procedural encoding might also constrain these truth-conditional inferential processes, and proposed that personal pronouns function in this way. In this thesis, I maintain a clear distinction between inferential processes which result in the construction of propositional conceptual representations, and inferential processes which take such representations as their input.

The main aim of this thesis is to determine what kinds of linguistic expressions encode procedural information constraining the construction of propositional conceptual representations, and to specify how they function. To this end, I shall discuss pronouns and grammatical markers of tense, aspect and modality. I shall conclude by demonstrating how a characterisation of grammatical markers as exponents of procedural encoding accounts for certain aspects of grammaticization and linguistic variation.

Since the arguments presented in this thesis are grounded in a general cognitive framework, they should be applicable cross-linguistically (assuming that the cognitive abilities of speakers of all languages are essentially the same). For this reason, I shall



make reference at various points in the thesis to Kiswahili, a Bantu language spoken widely in Tanzania, Kenya, Uganda and eastern Zaire.<sup>2</sup>

To conclude this introduction, I shall provide a justification for working within the framework of relevance theory (§0.2) and an overview of the thesis as a whole (§0.3).

## 0.2 Why relevance theory?

In this section I shall briefly address arguments in defence of the choice of relevance theory as the theoretical basis (and to some extent the object) of this thesis. I cannot hope to provide an exhaustive justification of relevance theory, merely arguments in support of the reasonableness of adopting this framework as opposed to certain others (as did Blakemore (1987) and Blass (1990) with respect to relevance theory and alternative theories such as Halliday & Hasan's (1976) Cohesion theory). I shall first indicate why the 'standard' Montague/Grice approach was not adopted, and then why relevance theory was selected in preference to other alternative pragmatico-semantic theories.

Traditional truth-conditional, model-theoretic semantics treats the content of linguistic expressions as functions from indices (possible-world/time pairs) to extensions. The semantic interpretation of a sentence is assumed to be directly truth-conditional relative to a given index, i.e. to a context. All aspects of interpretation which are not subject to a truth-theoretical analysis can be treated pragmatically, i.e. pragmatics = meaning - truth conditions (Gazdar 1979:2). Grice's Co-operative Principle and maxims and some version of speech act theory have traditionally been employed to deal with non-truth-conditional aspects of meaning.

Various arguments have been levelled against traditional Montague semantics<sup>3</sup> but I shall address just two of its basic assumptions. In stating that a sentence can be directly truth-conditionally determined relative to a given context, it is assumed a)

---

<sup>2</sup> Although Kiswahili (or Swahili) has been influenced by English, for example in the use of relative clause constructions by bilingual Kiswahili-English authors (Russell and Rajabu 1995), such influence is minimal.

<sup>3</sup> For example, Jardine (1975) is an early critique noting that it is easier to translate a formal language into natural language, say by introducing rules which selectively replace names and definite descriptions with appropriate pronouns, than it is to translate natural language into a formal language by assigning appropriate references to pronouns using contextual information (Jardine 1975:229); Le Pore (1983), Higginbotham (1988), Etchemendy (1988) and Bickhard & Campbell (1992) provide theoretical critiques of model theory, as does Kempson (1988b) with respect to truth-theoretical semantics. Hirst (1987:32) offers a practical criticism in the strong claim that "truth-conditional semantics is not useful in AI." Such criticisms also impinge on GPSG, a non-standard (non-GB) syntactic theory explicitly linked to Montague semantics (Gazdar *et al* 1985).

that contexts are given, and b) that sentences can be directly truth-conditionally determined. Both of these assumptions are without foundation. Throughout chapter 1 (and in particular in §1.1, 1.3 and 1.4) I shall argue that utterance interpretation contexts are psychological entities which are created rather than given, and which are dependent on the linguistic content of the utterance in question rather than being predetermined. In fact, the notion of procedural information which informs this study originated with the idea that there might be linguistically encoded information the sole purpose of which is to contribute to the psychological process of context formation. The assumption that sentences can be directly truth-conditionally determined is also false. Relevance theory claims that sentences constitute the input to a modular process of decoding which gives rise not to truth-functional propositions but to logical forms (representations in the language of thought - see §1.3.1). This is argued for in §1.2 and 1.3. The utterance interpretation process is triggered by linguistic input, giving rise to a logical form; the addressee then constructs an utterance interpretation context and develops the logical form into a propositional form (which can then be truth-theoretically interpreted) by a process of inferential enrichment (described in §1.3.2). If the decoding process has to be combined with inferential processes to yield propositions, then languages - that is, grammatical systems - underdetermine propositions and cannot be directly truth-theoretically determined. One of the aims of chapter 1 is to argue for and elaborate on this view.

The flip side of the above arguments is that the traditional view of pragmatics as exclusively non-truth-theoretical is no longer tenable. In §1.3.3 it will be shown that the distinction between semantics as that which is encoded and pragmatics as that which is inferred cuts across the truth-conditional/non-truth-conditional distinction.

Having presented just a few arguments against adopting a 'traditional' approach, I shall briefly discuss why it is reasonable to adopt relevance theory rather than any of the other alternative pragmatico-semantic theories currently on offer, in particular Situation Semantics (Barwise & Perry 1983) and Discourse Representation Theory (DRT) (Kamp 1981, Kamp & Reyle 1993). Firstly, it is still too early to be able to state that any one of these theories is superior to the others in terms of internal logical coherence, empirical corroboration, extensibility and simplicity. Moreover, competition between rival theories is healthy, or as Lakatos (1978:69) puts it, "'Theoretical pluralism' is better than 'theoretical monism'." This said, the choice of relevance theory over Situation Semantics and DRT is motivated in part by the following considerations. In Situation Semantics sentences denote language-independent 'situation types' rather than truth-values, and utterance interpretation involves the recognition of the relation between the situation described (by a

sentence) and the situation in which it is uttered. Situation Semantics recognises the underdeterminacy of linguistic content (what it calls the 'efficiency of language') - hence sentences do not denote truth values - but, in its current form, rejects the idea of mental representations. Meaning in Situation Semantics is the recognition of a relation between situation types rather than the construction of a mental representation. §1.2 provides arguments (not exhaustive, of course) for Fodor's modularity hypothesis which underlies relevance theory and which presupposes the existence of mental representations (as does Chomsky's theory of UG).

DRT attempts to provide a formal representation of the utterance interpretation process which takes into account the dynamic nature of context formation (as recognised within the tradition of 'update semantics', e.g. Heim's (1983) File Change Semantics). In line with Grice's (1989:25) assumption that the only context-dependent processes which contribute to the interpretation of what is explicitly (as opposed to implicitly) communicated are disambiguation and reference assignment, work within DRT has focused on quantification (since resolution of quantifier vagueness is a prerequisite to the construction of a Discourse Representation Structure) and nominal anaphora (and to a lesser extent temporal anaphora, e.g. Partee (1984)). However, DRT fails to provide a model of context formation which takes account of its essentially psychological nature (for example the ability to draw on encyclopaedic knowledge in the interpretation of bridging reference). DRT (as given in Kamp & Reyle (1993)) updates the interpretation context (represented by a Discourse Representation Structure) of a text purely on the basis of previous discourse, and does this incrementally, i.e. each new sentence is incorporated into the principle (most recent) DRS to form a new DRS. There are two drawbacks to this procedure: First, DRT has problems coping with multiple-sentence discourses, since anaphoric reference need not be to the most recent DRS; what DRT lacks is a mechanism or heuristic by which the most appropriate DRS for anaphora resolution can be selected. Second, the initial context (or 'starting DRS') is empty and the formation of each subsequent DRS results purely from the incorporation of the following sentence into the prior DRS. According to Kamp & Reyle (1993:85) "nothing of importance is lost" if the initial context is assumed to be empty, and for the purposes of DRT this may well be true, since DRT is an attempt to model, in an idealised fashion, the capabilities of natural language and to imitate this process. Relevance theory, on the other hand, attempts to explain how everyday human communication is possible and therefore such idealisation is inappropriate. In addition, DRT does not attempt to represent the process of inferential enrichment of semantically underdetermined expressions (discussed in §1.3.2), which is an integral aspect of the interpretation of explicit communication and is also context-dependent.

### 0.3 Overview

In discussing procedural encoding, a distinction will be made between procedural encoding which constrains the construction of propositional conceptual representations, and procedural encoding which constrains the processing of such conceptual representations once recovered. This distinction is couched in a model of the utterance interpretation process in which three distinct phases are recognised: a linguistic decoding phase, an inferential phase which results in propositional conceptual representations, and a further inferential phase which computes the relevance of these. In chapter 1 I shall introduce this three-phase model of utterance interpretation and discuss the importance of inferential processing and the notion of relevance.

In chapter 2 the conceptual/procedural distinction will be introduced within the context supplied in chapter 1. I shall discuss the theoretical basis of the distinction and provide an analysis of various exponents of procedural encoding which constrain the processing of propositional conceptual representations. In chapter 3 I shall extend the analysis to incorporate procedural encoding which constrains the construction of propositional conceptual representations, discussing pronouns, and tense and aspect markers. In chapter 4, the principles underlying the analysis of tense and aspect will be applied to a more detailed account of the English modal auxiliaries *can*, *may*, *must* and *should*, drawing on previous relevance theoretic accounts.

A truly comprehensive analysis of grammatical markers must take account of variations in distribution and interpretation, and should also be compatible with a model of grammaticization - the process whereby grammatical markers develop out of lexical expressions. In chapter 5 I shall relate grammaticization to the extended analysis of the conceptual/procedural distinction provided in the previous chapters. It will be shown how the conceptual/procedural distinction is compatible with research into grammaticization, and how it allows for monosemous synchronic accounts of grammatical markers capable of accommodating linguistic change and variation. In chapter 6 such accounts will be provided for the grammatical markers *will*, *shall* and *be going to*.

Chapter 7 summarises the main conclusions drawn in the thesis, and their implications.

# **1 RELEVANCE THEORY: AN INTRODUCTION**

The aim of the current chapter is to introduce the central tenets of relevance theory and to lay the foundation for the conceptual/procedural distinction to be introduced in chapter 2. I shall do this within the framework provided by the three-phase model of utterance interpretation to be introduced in §1.1. This model makes a distinction between coded and inferential communication which is motivated by the modularity hypothesis (Fodor 1983). In §1.2 I shall outline the main features of the modularity hypothesis and review some of the experimental psycholinguistic evidence pertaining to it.

§1.3 and §1.4 deal with inferential communication. In §1.3 I shall discuss the relevance theoretic account of the role of inference in the identification of propositions, and compare this with the traditional Gricean approach. In §1.4 I shall discuss the role of inference once a proposition has been recovered, and introduce the notion of relevance. It is these inferential phases of utterance interpretation which are constrained by procedural information; I shall therefore begin by emphasising the centrality of inferential (as opposed to coded) communication within relevance theory.

## **1.1 Inferential Communication**

### **1.1.0 Introduction**

Although both conceptual and procedural information types can be linguistically encoded, the conceptual/procedural distinction itself functions during inferential rather than coded communication. The significance of the conceptual/procedural distinction in linguistic semantics is therefore proportionate to the relative importance of inferential communication in the utterance interpretation process; in §1.1.1 I shall summarise the arguments underlying the relevance theoretic claim that inferential communication takes precedence over coded communication, a claim addressed in further detail in §1.3. In §1.1.2 I shall introduce a three-phase model of utterance interpretation in which one coded and two inferential phases are distinguished. This three-phase model of the utterance interpretation process informs not only the current chapter, but the thesis as a whole.

### 1.1.1 The primacy of inferential over coded communication

Meaning can be characterised either extensionally, as a relation between an external sign and an objective referent in the world, or intentionally, as a relation between a sign and a representation in the mind. The former characterisation is associated with the semiotic approach, and following Saussure has been the dominant paradigm within linguistics. Relevance theory, on the other hand, adopts the latter approach which dominates in cognitive science and artificial intelligence.

The radically cognitive nature of relevance theory is evident from Sperber & Wilson's characterisation of contexts as psychological constructs, being subsets of hearers' assumptions about the world (S&W 1986:15). Since psychological constructs are internal to individuals and not directly amenable to observation by others, contexts as defined by Sperber & Wilson are uniquely determined for each participant in a conversation. Therefore, it is impossible to talk of *the* (unique) context of an utterance (or of *the* (unique) interpretation of an utterance), rather, each context is determined relative to an individual and is accessible to that individual alone. This entails that mutual knowledge can play no role in communication. Mutual knowledge is *psychologically* unfeasible, since, by definition, people who share mutual knowledge must know for certain that they share this knowledge; given what has been said about contexts above, such certainty is an ideal rather than a psychological reality, (cf. Smith 1982 for further discussion of mutual knowledge). Since communication is a psychological process, mutual knowledge can play no part in it, hence the definition of a psychological construct as a subset of *assumptions* rather than of knowledge.

The rejection of mutual knowledge as a basis for utterance interpretation leads Sperber & Wilson (S&W 1986:18) to reject the traditional (Saussurian) code model of communication; if inference has no role to play in verbal communication, that is, if a coding-decoding process alone is sufficient, then communication is only possible where every coded element has a single, unambiguous referent. Such a situation could only occur within a strictly determined, wholly shared context for which mutual knowledge is a necessity. This argument is developed in line with proposals made by Grice, whose most original contribution to the study of communication, according to Sperber & Wilson (1986:25), was to suggest that communication is possible even in the absence of a code. It has been claimed (e.g. Stevenson (1993:136), Vallduví (1993:32)) that relevance theory is basically a modification of Grice's proposals; there is a sense in which this is the case, but relevance theory is more than simply a reduction of the Co-operative Principle and maxims to the single maxim of relation, recast as the principle of relevance (see §1.4.2). The existence and use of determinate

linguistic signs in communication is not denied in relevance theoretic work - on the contrary, the accurate characterisation of such signs and the ways in which they are used in communication is a central concern of this thesis. Nonetheless, in relevance theory, the coding-decoding process is seen as subservient to the inferential process of the Gricean model, according to which communication involves the production and interpretation of evidence (S&W 1986:27). This subservience of the coded system to inference is fundamental to relevance theory.

Having rejected mutual knowledge as a basis from which inferences can be drawn, Sperber & Wilson introduce the weaker notion of *mutual manifestness* (1986:39-45). Any assumption constructed by an individual is 'manifest' if it can be accepted as being true or probably true by her; it is therefore possible for there to be degrees of manifestness, and for an assumption to be both manifest and false. Mutual manifestness need not, therefore, involve certainty, and so is psychologically feasible. A set of manifest assumptions constitutes an individual's *cognitive environment*. To illustrate these notions Sperber & Wilson give the following example:

“Suppose Peter and Mary are looking at a landscape where she has noticed a distant church. She says to him,

(49) I've been inside that church.

She does not stop to ask herself whether he has noticed the building, and whether he assumes she has noticed, and assumes she has noticed he has noticed, and so on, or whether he has assumed it is a church, and assumes she assumes it is, and so on. All she needs is reasonable confidence that he will be able to identify the building as a church when required to: in other words, that a certain assumption will be manifest in his cognitive environment at the right time.”

(S&W 1986:43-44)

With mutual manifestness rather than mutual knowledge as a basis, communication must be assumed to be governed by a less than perfect heuristic rather than a fail-safe mechanism. Occasional failures in communication are therefore to be expected; what requires explanation is the fact that communication is so often successful. The notion of relevance is proposed as an explanation of this fact: successful communication makes manifest the speaker's informative intention, "to make manifest or more manifest to the audience a set of assumptions {I}." (S&W 1986:58). Before addressing relevance itself, I shall propose a model of the utterance interpretation process within which the notion of relevance will be situated; I shall then discuss the roles of coded and inferential communication within this model.

### 1.1.2 Three phases of utterance interpretation

Broadly speaking, utterance interpretation is viewed in relevance theory as a two-tier system consisting of decoding and inference. In this thesis I shall recognise inference as having two distinct functions: the identification of the proposition expressed and the identification of intended contextual effects. Thus, utterance interpretation as a whole can be viewed as a three-phase process; for convenience I shall speak of *phase one*, *phase two* and *phase three* of utterance interpretation, as follows:

#### *Phase One*

A linguistically encoded utterance is decoded to give an abstract semantic representation, or (possibly incomplete) **logical form**.<sup>4</sup>

#### *Phase Two*

The **logical form** is inferentially, or contextually enriched resulting in the identification of a **propositional form**.<sup>5</sup> A propositional form is a **conceptual representation** which may be propositional, corresponding to a fully understood idea; or, a propositional form may be what Sperber (1985:51) calls a 'semi-propositional representation', which fails to identify a unique proposition, and corresponds to a half-understood idea.

#### *Phase Three*

The **conceptual representation** is combined with other conceptual representations (prior assumptions, background knowledge, etc.) in order to achieve **contextual effects**.

Although this three-phase model is not explicitly employed in other relevance theoretic work, it is implicit in the theory as it derives, as will be seen, from basic relevance theoretic assumptions about the utterance interpretation process. Figure 1 (below) provides a preliminary sketch of the model of utterance interpretation to be developed:

---

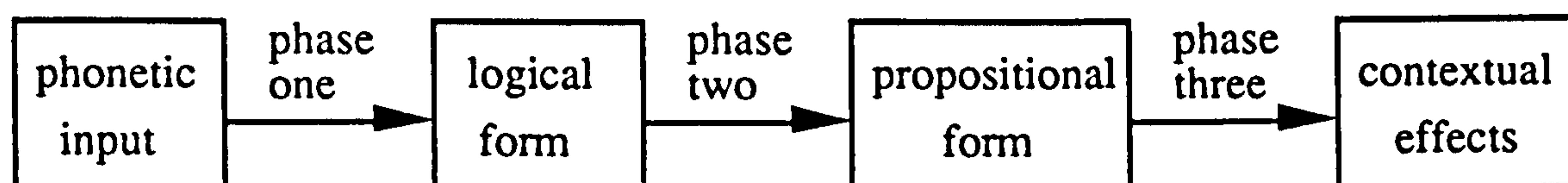
<sup>4</sup> See §1.3.1 below for a discussion of logical form.

<sup>5</sup> A *propositional form* has an associated *propositional attitude* (expressing, for example, illocutionary force) which aids in the recovery of contextual effects; for example, (1a) and (1b) have different propositional attitudes but the same propositional form, for example that in (2):

- |     |   |   |                             |
|-----|---|---|-----------------------------|
| (1) | a | Mary: You shouldn't.                        |                             |
|     | b | Peter: Shouldn't I?                         |                             |
| (2) |   | Peter shouldn't smoke when Mary is present. | (Wilson & Sperber 1988:134) |



Figure 1: The three phase model of utterance interpretation



What figure 1 fails to indicate, is that the three phases are neither wholly autonomous nor wholly sequential. For example, inferential processes can inhibit the first phase of interpretation (construction of a logical form), and can do so at the level of individual words as well as of larger units of discourse (see §1.3.1 below), and the search for adequate contextual effects can play a decisive role in the identification of a propositional form, say, by contributing to the determination of reference assignment by pronouns (discussed in §3.1). This is ultimately a result of the fact that utterance interpretation is relevance-driven; every act of ostensive communication is assumed to be optimally relevant, a notion which will be addressed in §1.4. This said, in the subsequent discussion of procedural encoding (that is, linguistically encoded expressions which constrain inferential processes) it will be necessary to distinguish between expressions which constrain phase three of utterance interpretation (the identification of the intended contextual effects of an utterance) and expressions which constrain phase two of utterance interpretation (the identification of a propositional form). For this reason I shall maintain a three-way rather than the more general two-way distinction (between decoding and inference); the three phases of utterance interpretation will be discussed in turn.

## **1.2 Phase One: Decoding**

### **1.2.0 Introduction**

I cannot hope to provide a comprehensive account of linguistic decoding in the short space available here; my intention is simply to summarise the relevance theoretic position, such as it is, on the role of linguistic coding in utterance interpretation and its interaction with the inferential processes which the theory and this study are primarily concerned to explain. This entails addressing Fodor's (1983) modularity hypothesis, as this provides the justification for recognising what I have termed phase one of utterance interpretation as distinct from the inferential phases. The modularity hypothesis will be introduced in §1.2.1, and the extent to which decoding determines the identification of logical form will be discussed in §1.2.2. I shall then (§1.2.3) assess the modularity preserving model of syntactic parsing which most closely corresponds to my findings thus far (the parallel weakly interactive model) before summarising and discussing the findings (§1.2.4).

### 1.2.1 The modularity hypothesis

The fundamental characteristic of the first phase of utterance interpretation (and production), is that it is modular. Following Fodor (1983)<sup>6</sup>, Sperber & Wilson (1986:71) assume that the mind consists of, on the one hand, input systems, which process various kinds of perceptual information (visual, auditory, etc.), and on the other hand, the central systems, which integrate information from the various input systems and from memory, and perform inferential tasks. The input systems are modules, which exhibit the following main characteristics:

1. Input systems are mandatory - the operation of each one is automatic.
2. They are fast - in 'close shadowing' experiments, where subjects have to repeat utterances as they hear them, the time lag between hearing and producing speech is typically less than 250ms.
3. They are domain specific - the processes of each module are unique to it and not shared with any other module.
4. Modules are cognitively impenetrable - that is, only the end product of modular processing is available to the central systems.
5. They are informationally encapsulated - they are unaffected by the beliefs, values, etc. of the central systems.

There is considerable experimental support for modularity, such as neuropsychological data from aphasia which shows that there is a language module which can be impaired in people who are otherwise not mentally impaired; analysis of grammaticality judgements by Broca's aphasics further demonstrates that "the linguistic processing system is itself modular, as shown by the divisions between lexical processing, syntactic parsing and the semantic interpretation of syntactic structures" (Bánréti 1994:29).

As for the central systems, Fodor (1983:140) claims that, given their informationally unencapsulated nature, no account can be given of them; that is, a psychology of the central processor would have to account for the potentially infinite number of factors which may have some bearing on the fixation of beliefs. (This is the essence of Fodor's 'First Law of the Non-existence of Cognitive Science'.) However, Sperber & Wilson (1986:66, 75) (and Wilson & Sperber 1986a) claim that an account of the relatively local inferential processes of utterance interpretation offers an insight into

---

<sup>6</sup>Although early formulations of relevance theory (e.g. Wilson & Sperber 1981) preceded Fodor (1983), the modularity hypothesis is part of the cognitive psychological framework upon which relevance theory is currently based. The idea of modularity was current prior to Fodor however; for example Chomsky (1980: 89) invoked a form of modularity: "I am tentatively assuming the mind to be modular in structure, a system of interacting subsystems that have their own special properties."

more complex functions of the central systems, such as scientific theorising - Fodor's paradigm case of a central process. Aside from these differences and differences over the nature and status of the central systems (Fodor assumes these are essentially inductive, Sperber & Wilson deductive), Sperber & Wilson follow Fodor in distinguishing between a language module and a non-modular central processor. More recently, however, Sperber (1994) has suggested that the central systems might be less modular rather than simply non-modular. Sperber (1994:49) suggests that the presence of certain concepts in a conceptual representation might activate specific central modules in a similar way to the activation of certain input modules by the presence of particular stimuli. Moreover, Sperber (1994:60) posits a meta-representational module which takes "concepts of concepts and representations of representations" as input, rather than simply first order concepts and representations of things, thereby facilitating communication (which is the processing of other people's beliefs and representations).

Returning to the input modules, both Fodor and Sperber & Wilson (1986) argue that because we cannot help but construct a semantic representation of any utterance we hear in a language known to us, even accidentally or unwillingly, linguistic decoding must be modular. Relevance theory has little to say concerning the decoding phase of utterance interpretation - this is properly the domain of phonological, morphological and syntactic theories - however, what detail is provided has some bearing on the more fully articulated inferential interpretation process. In this section I shall be concerned primarily with the interaction between the modular decoding phase of interpretation and the central inferential processes.

### 1.2.2 The role of linguistic decoding in the identification of logical form

Identifying the intended logical form of an utterance is not done simply by mechanically applying linguistic rules. Evidence from (minimally contextualised) garden path utterances suggests that the outcome of any such decontextualised disambiguation procedure is rejected if it is inconsistent with some non-syntactic criterion (see examples below). According to Sperber & Wilson (1986:184) this, ultimately, is the criterion of consistency with the principle of relevance (described in §1.4). The identification of logical form therefore involves both a linguistic input module and a central inferential process. It is the relationship between these that is the concern of the current section. A common view (assumed in the models described below) holds that grammar is modular and that its form and functions can (and indeed must) be described independently of considerations pertaining to the central systems. What is made accessible to the central systems consists of concepts associated with

individual words and a syntactic structure. An alternative view holds that "the structure of language is explained by basic principles of the general cognitive system as a whole" (Groefsema 1992:276). On this latter view, which is still modular, the linguistic processor consists (in addition to a prosodic processor) of a mental lexicon containing phonological and orthographic forms, the output of which is a string of concepts accessible to the central systems. Groefsema (1992) does not claim that there is no syntactic structure, merely that syntactic structure is derivative on logical structure, and that no independent syntactic structure is created during utterance interpretation. Experimental evidence from Bock & Loebell (1990) which demonstrated that speakers tended to produce sentences with structural (syntactic) forms similar to those of topically unrelated priming sentences (cf. Bock *et al* 1992) suggests that syntactic structures are created and stored during utterance interpretation, but this does not determine whether syntactic structure is generated independently of logical structure or is derivative on it.

Whether an utterance gives rise to a set of concepts plus a syntactic representation, or to concepts only, the question of exponency, that is, of the relation between the two levels (linguistic and cognitive), remains to be resolved. There are two basic choices to be made. First is the choice between a parallel processing model, in which the input module presents the central systems with a choice of semantic representations, and a serial model, in which the input module delivers one representation at a time, beginning with the easiest to construct and only constructing the second easiest if the first is rejected, and so on. Second, there is a choice over the point at which semantic representations are made available to the central systems; in a highly modular model with a minimum of interaction between the modules and the central systems this point would be as late as possible and would include a syntactic representation, but in a weakly modular model semantic representations could be available earlier, say, at the end of each word, in which case a syntactic representation could, but need not, be represented. With regard to these alternatives, Sperber & Wilson (1986:206) assume that the central systems perform computations of semantic interpretations of syntactic category labels, thus implying that a syntactic representation is part of the output of a language module. Elsewhere, they suggest that:

"the input module might construct all the linguistically possible interpretations of the first constituent of the sentence, and submit them to the central mechanism, which would, when possible, choose one of them and inform the linguistic module of its choice. As a result, the module's decoding processes would be partly inhibited."

(S&W 1986:186)

Although Sperber & Wilson's model is, as they admit, purely speculative and non-technical, it is sufficiently precise to be compared with other more comprehensive

models. Modularity is clearly maintained since the module remains informationally encapsulated, having no access to encyclopaedic contextual information; modular processes *are* affected by contextual information, but only in an inhibitory way. It is also clearly a parallel processing model, and one in which linguistic interpretations are made available earlier, at the end of a constituent, rather than later.

Relevance theory is a partial theory of performance in that it provides a model of the inferential phases of utterance interpretation; if this is to contribute to a complete model of interpretation, relevance theory must be compatible with a defensible model of parsing. Such a model must allow clear delimitation of the inferential phases of interpretation from linguistic decoding in theory. Stevenson (1993:79) lists three basic models of parsing which assume modularity: serial independent (Frazier 1979), parallel weakly interactive (Crain & Steedman 1985, Altmann & Steedman 1988) and connectionist (Walz & Pollack 1985). These differ primarily over whether or not a syntactic representation is represented and the point at which semantic and pragmatic processes are introduced. I shall briefly compare each type of model with Sperber and Wilson's tentative proposals.

Frazier's (1979) serial independent model is extremely modular in that it assumes syntax and semantics to be autonomous in practice as well as in theory. Cases of syntactic ambiguity are therefore resolved initially by purely syntactic strategies, such as 'late closure' and 'minimal attachment' which state that the syntactic parser will automatically attach incoming material to material on its left which has already been analysed (Frazier 1979:114). Only if the resulting semantic interpretation of the whole clause is incompatible with the context will the processor backtrack and reparse the sentence. This model is obviously inconsistent with Sperber & Wilson's model of a parallel syntactic parser which makes the various interpretations of each constituent immediately available to the central systems.

Closer to Sperber & Wilson's view is the parallel weakly interactive model of Crain & Steedman (1985) and Altmann & Steedman (1988). As evidence against the serial independent model, Crain & Steedman (1985) noted that sentences such as (1.1):

(1.1)           He told the woman that he was having trouble with to leave.

are garden path sentences when presented out of context (that is, as isolated written examples), a complement reading being the preferred interpretation. However, given the preceding context clause in (1.2), the relative clause interpretation is not only easier to construct, but seems to be the preferred reading:

- (1.2) The psychologist saw two women. He told the woman that he was having trouble with to leave.

Their experiments suggested that: 1: referential context can influence garden path effects; 2: it can do so even while the clause remains incomplete; 3: residual effects, such as those predicted by a serial model, are not apparent.<sup>7</sup>

This led to the development of a weakly interactive model of parallel syntactic parsing. Crain & Steedman (1985:325) distinguish between weakly interactive models, in which syntactic processing proposes alternative interpretations which inferential processing selects or rejects, and strongly interactive models, in which semantics and context influence which syntactic entities get proposed in the first place. Weakly interactive models are modularity preserving but strongly interactive models are not, since central processes can give the syntactic parser feedback about the contextual acceptability of its output (Fodor 1983:134-5), but semantic or pragmatic information cannot be used predictively to guide linguistic parsing. The strong hypothesis is untenable because, according to Altmann & Steedman (1988:206), the context could only produce such specific effects if there were already an interpretation available, but the only way that such detailed interpretations can be obtained is via syntax. Similarly, their preference for parallel over serial processing rests on the theoretical implausibility of a serial model once weak interaction is assumed:

"We can only reject [an interpretation] in comparison with some more plausible alternative, simply because it may be the *only* analysis. We therefore claim that weakly interactive processors must by definition propose syntactic alternatives for semantic and pragmatic adjudication *in parallel*."

(Altmann & Steedman 1988:208)

However, if we also assume that weak interaction occurs as soon as constituents are identified, and that these are smaller (e.g. words) rather than larger (e.g. clauses or sentences) - what Altmann & Steedman (1988) term "fine-grained" processing - then serial and parallel processing will be empirically indistinguishable.

---

<sup>7</sup> Eye movement experiments conducted by Altmann *et al* (1992) support these findings. Whilst the serial independent model would account for the preferred reading of sentences like (1.1) on the grounds of minimal attachment, Altmann *et al* (1992) propose a cognitive cause. The context presupposed by the use of the linguistic form "the woman" (a universe of discourse containing one accessible instance of an adult female) is more easily constructed than the context presupposed by the identifying clause "the woman that he was having trouble with" (a universe of discourse containing more than one woman).

Whereas in the parallel weakly interactive model described above, the fine-grained nature of the parser renders the distinction between parallel and serial processing of little practical importance, it is central to the connectionist model of Walz & Pollack (1985) (cf. Tanenhaus *et al* 1987). This is a massively parallel system in which syntactic, semantic and contextual components analyse the input in parallel. Because syntactic units are subject to excitation or inhibition from the other components only after they have been activated, the connectionist model maintains modularity and can be viewed as weakly interactive. However, Walz & Pollack's connectionist model is incompatible with Sperber & Wilson's because it posits excitatory links in addition to inhibitory ones, and, more importantly, it assumes a modular structure for *pragmatic* processes (Stevenson 1993:80). Wilson & Sperber (1986a:67) explicitly deny that pragmatics is modular, as this would amount to saying that there is a pragmatic code - contrary to the relevance theoretic claim for the independence of ostensive-inferential communication from coded communication.<sup>8</sup>

### 1.2.3 Assessing the weakly interactive model

The closest model to Sperber & Wilson's suggestions regarding the processes underlying the recovery of propositional form would appear to be the 'fine-grained' parallel weakly interactive model. Two questions now need to be addressed: How plausible is this model? and What are its implications for relevance theory? Support for the parallel weakly interactive model comes from experiments conducted by e.g. MacDonald (1993), Ni *et al* (1996), Trueswell & Tanenhaus (1991), and Trueswell *et al* (1994), which suggest that ambiguity in the processing of information in one domain (say, syntactic information or semantic information) can be constrained by transferring processed information from other relevant domains. On the other hand, criticisms have recently been made of the weakly interactive model from both a serial independent perspective (Nicol & Pickering 1993) and a strongly interactive perspective (Marslen-Wilson *et al* 1993).

In their experiments, Nicol & Pickering (1993) showed that a relative clause interpretation (1.4) is still a viable option at the point of the embedded verb 'phoned', at which point, they claim, only the complement clause interpretation (1.3) ought to be viable in a weakly interactive model:

---

<sup>8</sup> As mentioned above in §1.2.1 (p.19), however, Dan Sperber (Sperber 1994) has suggested that pragmatic processing might involve meta-representational skills which are modular, that is, that there might be modules which take representations made available by other modules as their input. This suggestion is still the object of considerable debate, and so will not be pursued further here.

- (1.3) The receptionist informed the doctor [that the journalist had phoned about the events].
- (1.4) The receptionist informed [the doctor that the journalist had phoned] about the events.

"the weakly interactive model assumes that both analyses are active momentarily, but that, without discourse support, the relative clause reading is rapidly rejected. Hence [it] would predict that, by the point at which the embedded verb is encountered, the complement clause analysis is the only one still under consideration."

(Nicol & Pickering 1993:226)

However this criticism is multiply flawed. First, the weakly interactive model posits *inhibitory* interaction with discourse, rather than "discourse support", so in the absence of contextual counter evidence both analyses will continue to be viable, as Nicol & Pickering in fact found. Second, Nicol & Pickering (1993:226) admit that the sentences "were deliberately intoned in such a way as to favor the relative clause analysis", but dismiss this variable because "these differences are extremely subtle at the point when [Crain & Steedman] predict a choice should be made"; however, Marslen-Wilson *et al* (1992) found that prosodic information aids syntactic disambiguation as does contextual support. Third, Nicol & Pickering (1993) failed to meet the criterion suggested by Altmann & Steedman (1988:204) that:

"any experiment that attempts to falsify the interactive model cannot depend upon the supposed neutrality of any contexts: Contexts must be actively manipulated to explicitly support or deny the relevant presuppositions."

Both Altmann & Steedman (1988) and Sperber & Wilson (1986) recognise that there is no such thing as a completely decontextualised sentence. Hearers will bring their encyclopaedic knowledge to bear on any sentence they encounter, as this, as well as previous discourse and the physical environment, is part of context:

"In the traditional psycholinguistic experiment, in which the sentences are presented out of context, definite expressions still carry their presuppositions<sup>9</sup>, which must be supported or accommodated by the hearer. In a case of local ambiguity it is reasonable to assume that the reading carrying fewest unsupported presuppositions, and therefore requiring fewest modifications to the database, will be favoured..."

(Altmann & Steedman 1988:203)

---

<sup>9</sup> I interpret Altmann & Steedman's use of the term 'presupposition' to refer to presuppositional effects rather than to presupposition as a technical linguistic term in opposition to focus. Sperber & Wilson (1986:217), whilst accepting the existence of presuppositional effects, reject the theoretical focus-presupposition distinction in favour of a foreground-background distinction which "is simply a descriptive label used to distinguish two complementary and independently necessary aspects of the interpretation process".



Thus, the complement clause interpretation (1.3) is to be favoured since it assumes fewer presuppositions than the relative clause interpretation (1.4) in which it is presupposed that the journalist had already phoned the doctor. However, if the complement clause interpretation is "favoured" this does not entail that the relative clause interpretation is rejected, which is what Nicol & Pickering (1993:226) claim the weakly interactive model predicts. Thus the parallel weakly interactive model has not been falsified or even weakened by Nicol & Pickering's supposed counter-evidence.

Arguing against the modularity hypothesis, Marslen-Wilson & Tyler (1987) suggested that processing which results in the recovery of discourse models (that is, processing which takes contextual factors into account) exhibits the same characteristics with respect to modularity as does processing resulting in a logical form. That is, they claim that processes which map onto discourse models are domain specific, mandatory, fast, and have limited access to intermediate representations - all typical input system characteristics. With regard to informational encapsulation Marslen-Wilson & Tyler (1987:41) note that the modularity hypothesis can be defeated if it is shown that central and modular processes are not isolated from each other in the ways required by the theory. This means that the crucial question is when semantic and pragmatic information interact with the syntactic parser. If the modularity hypothesis is correct, this should be after the first pass syntactic output becomes available to central processes.

The aim of the four sets of experiments reported in Marslen-Wilson & Tyler (1987) is to demonstrate that pragmatic considerations can influence utterance interpretation *prior* to syntactic processing. In one of these types of experiment, reported in greater detail in Marslen-Wilson *et al* (1993), subjects heard short stories ending in an incomplete fragment. The discourse focus, type of anaphor, and pragmatic inference in the stories were systematically varied. At the end of the incomplete fragment subjects were given a visual probe and the time it took them to name this was measured. It was found that, whatever the nature of the subject anaphor or focus, the probe was invariably checked for pragmatic coherence with listeners' presuppositions, suggesting to Marslen-Wilson *et al* that pragmatic inference can determine reference assignment in the absence of support from discourse factors, and that inference takes precedence over focus when these conflict. They conclude that this analysis favours a strongly interactive (and hence non-modular) model of syntactic processing.

This conclusion is only valid if the output of syntactic parsing is 'course-grained', that is, if only complete syntactic outputs (whole clauses) are made available to the central processes. Given the fine-grained account proposed in Altmann & Steedman (1988), however, inferential processes can affect first-pass syntactic parses in the way described in Marslen-Wilson *et al* (1993) without violating modularity. Marslen-Wilson & Tyler (1987:51) admit that "no matter how early context effects are detected, it is always possible to argue that multiple readings were nonetheless computed, so that what we are picking up are after-the-event selection effects rather than direct control of the initial syntactic parse." This, they claim, is nonetheless bad for the modularity hypothesis since it entails that modular accounts are empirically indistinguishable from non-modular accounts. The two are indeed empirically indistinguishable, but this is as bad for non-modular accounts as for modular ones. The only implication of this observation for the modularity hypothesis, is that informational encapsulation becomes a theoretical constraint on modules rather than an empirically observable heuristic for determining what is or is not a module.

In conclusion, two key observations were made in Marslen-Wilson & Tyler (1987:43): (i) that the mapping from an incoming utterance onto a discourse model is as rapid as the mapping onto a logical form; (ii) that the discourse mapping process is not significantly slowed down by pragmatic inferences. This, they claim, provides clear evidence for top-down influences on syntactic choice during first-pass syntactic processing. If this is the case, informational encapsulation has been violated, and the modularity hypothesis becomes untenable. Against this view I have argued that a fine-grained, incremental approach (such as the parallel, weakly interactive model discussed above) can accommodate observations (i) and (ii), since the construction of a discourse model takes place on the basis of partial information before complete syntactic structures have been made available to the central processes.

#### 1.2.4 Summary

The account of linguistic (as opposed to inferential) processing with which relevance theory is most clearly compatible is the fine-grained, parallel, weakly interactive model of Crain & Steedman (1985) and Altmann & Steedman (1988). This account is modularity preserving, but empirically indistinguishable from strongly interactive accounts such as that of Marslen-Wilson & Tyler (1987) and Marslen-Wilson *et al* (1993).

According to relevance theory, the output of linguistic decoding is a (possibly incomplete) logical form, which is completed and inferentially enriched by the central

inferential system to correctly identify the propositional form of the speaker's utterance. Put another way, inferential enrichment of logical form leads to the construction of a mental representation in the hearer's mind, corresponding (if successful) to the intended propositional form of the utterance. In practice, assuming that a) utterance interpretation is fine-grained - i.e. that linguistic analyses of input are made accessible to the central processes as soon as possible, and b) that the central processes begin interpreting the output of the language module immediately, an entire, unenriched logical form should rarely, if ever, be entertained in the mind of an addressee. In theory, however, it is possible to talk of complete, unenriched logical forms as distinct from propositional forms. Bearing this in mind, I shall now discuss how logical forms are inferentially enriched.

### **1.3 Phase Two: Inferential Enrichment**

#### **1.3.0 Introduction**

In this section, I shall provide a brief characterisation of logical form (§1.3.1) followed (in §1.3.2) by a non-technical account of how logical form is inferentially enriched to recover the correct propositional form.<sup>10</sup> In its defence, I shall compare the relevance theoretic approach with that of Gricean pragmatics (§1.3.3).

#### **1.3.1 Logical form**

Logical forms are well-formed formulae in the internal language of thought; although derived from linguistic input, they are not themselves linguistic representations. Sperber & Wilson's use of the term 'logical form' therefore differs from Chomsky's:

“On GB assumptions, the LF of a sentence is a construct of the same type as the S-structure from which it is derived, one of a set of indexed, labelled bracketings which jointly constitute the syntactic analysis of the sentence. By contrast, the logical form of a sentence on relevance-theoretic assumptions is an incomplete expression of a different formal system, that of the internalised language of thought, an inferentially transparent system whose completed formulae display all the properties of formal systems devised to reconstruct inference - no ambiguity, no indexical elements.”

(Kempson 1988b:20)

---

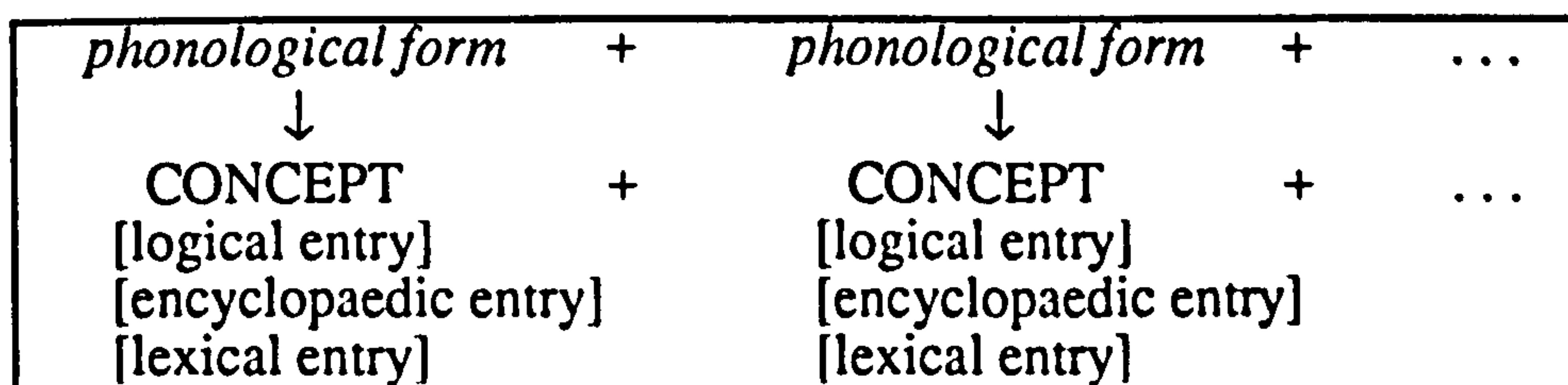
<sup>10</sup> A technical model in the form of a Labelled Deductive System has been proposed by Ruth Kempson (e.g. Kempson 1994a, 1994b); in this, inferential enrichment has been "modelled as a reasoning task over logical type specifications driven by lexical input, with additional choices being made by the hearer in all cases where the input does not fully dictate the selected interpretation" (Kempson 1994b:144-5).

A logical form, as characterised here, is an abstract semantic representation consisting of a well-formed string of concepts. Following Sperber & Wilson (1986:86), I assume that concepts are abstract psychological labels, or addresses, containing logical, encyclopaedic and lexical entries, defined as follows:

"The *logical entry* for a concept consists of a set of deductive rules which apply to logical forms of which that concept is a constituent. The *encyclopaedic entry* contains information about the extension and/or denotation of the concept.... The *lexical entry* contains information about the natural-language counterpart of the concept..."  
(S&W 1986:86)

This is represented in a simplified form<sup>11</sup> in figure 2

Figure 2: Concepts in a logical form



A word gives rise to a concept containing a logical entry which helps determine the meaning of the word when it is used in conjunction with other words in an utterance; I shall postpone discussion of the nature of the logical entry until §2.2.3, in which Groefsema's (1992) characterisation of logical entries in terms of 'logical selection frames' will be discussed. The encyclopaedic entry specifies the assumptions held about the extension and/or denotation of the concept, and is not objectively determined, but will vary from individual to individual; the encyclopaedic entry for a given word may differ between speaker and hearer since each interlocutor's encyclopaedic data base is unique, being the result of numerous and diverse experiences. The encyclopaedic entry is also typically semantically incomplete or underdetermined (see below). What Sperber & Wilson (1986:86) call the lexical entry specifies the phonological structure, syntactic category membership and syntactic co-occurrence restrictions (presumably strict subcategorization rules etc.) of the lexical counterpart of the concept. There is some debate as to whether a lexical entry should be part of a concept; Groefsema (1992:155) argues that the only relevant information in the lexical entry of a concept, so far as processing is concerned, is its phonological structure, since syntactic structure is, she argues, derivative on logical structure. As I mentioned in §1.2.2, the relation between syntax and logical structure is far from

<sup>11</sup> Logical forms are far more complex than figure two suggests as, for example, Jackendoff (1983) demonstrates.

clear, however syntactic structure (whether independently motivated or not) is created during processing, as demonstrated by Bock & Loebell (1990). Since processing is incremental, operating on incomplete syntactic structures, I assume that any syntactic information contained in the lexical entry is akin to the incrementally sensitive syntactic categories of, for example, left-associative grammar (Hausser 1989), although no specific syntactic theory or approach is assumed.<sup>12</sup>

### 1.3.2 Disambiguation, reference assignment and inferential enrichment

In the previous section a logical form was defined as a well-formed string of concepts, each of which gives access to the information in its logical, encyclopaedic and (possibly) lexical entries. The information provided by the logical (and lexical) entries of the concepts in a logical form, being necessary for the construction of the logical form, is immediately and fully accessible to an addressee. In contrast, the information provided by the encyclopaedic entries of the concepts in a logical form is not accessed in its entirety; the subset of encyclopaedic information used in the interpretation of an utterance is restricted to information which has been made immediately accessible, through the construction of an utterance interpretation context.

The specification of this subset of encyclopaedic information is a necessary step in the development of a logical form into a fully propositional form. The process involved in this development is that of inferential enrichment:

“Linguistically encoded semantic representations [logical forms] are abstract mental structures which must be inferentially enriched before they can be taken to represent anything of interest.”

(S&W 1986:174)

For a logical form to represent anything of interest, in the sense of the above quotation, it must be capable of achieving adequate contextual effects (see §1.4). This does not necessarily mean that a logical form has to give rise to a fully propositional representation, corresponding to a fully understood idea to be of interest; as Sperber (1985) notes, half-understood ideas (including religious, philosophical and often scientific beliefs) entertained as ‘semi-propositional representations’ may also be relevant. (All the cases discussed in this thesis, however, involve propositional representations.)

---

<sup>12</sup> Left-associative grammar is mentioned because the process of inferential enrichment to be discussed below requires that sentences typically be underdetermined in terms of both the lexical semantics of their constituent words and of their syntax; Hausser (1989) posits a suitably underdetermined syntax and a correspondingly important role for pragmatics.

Sperber & Wilson (1986:183-193) distinguish three inferential processes which a logical form may undergo in order to represent something of interest to an addressee: reference assignment, disambiguation, resolution of vagueness and the recovery of ellipsed material. The application of these three processes can be illustrated in sentence (1.5), the intended propositional form of which is unclear out of context:

(1.5) Peter's bat is too grey. [amalgam of S&W (1986) (20) and (21)]

When (1.5) is uttered, a referent must be assigned to *Peter*, although *Peter* is not ambiguous, reference assignment must nonetheless distinguish the intended Peter from various other Peters. *Bat*, on the other hand, is truly ambiguous; syntactic decoding will be able to distinguish *bat* as a verb from *bat* as a noun, but the noun *bat* is still many ways ambiguous; pragmatic disambiguation disposes amongst alternative meanings of *bat* and, if successful, ensures that a unique concept is recovered.

Resolution of vagueness resolves what Sperber & Wilson (1986:188) call the *semantic incompleteness*, rather than the ambiguity (that is, polysemacity) of the genitive encoded in "Peter's". If *bat* refers to a flying mammal, "Peter's bat" might indicate the bat owned, chosen, killed, discovered, mentioned, seen, etc. by Peter. The potential number of relations covered by the genitive is infinite, yet specific relations are typically understood in actual utterances. (In some sentences, only one interpretation of the genitive is possible: for example, in "The Huns' destruction of the city" (Napoli 1993:158) the only possible interpretation is that the destruction of the city was done by the Huns.) Sperber & Wilson (1986:188) therefore conclude that "contextual information is needed to resolve what should be seen as the semantic incompleteness, rather than the ambiguity of the genitive."

Sperber & Wilson (1986:188) claim that *too* is also semantically incomplete since not only can *too* modify a whole range of adverbs and adjectives, each in various ways, but in order to answer the question 'too grey *for what?*' in relation to (1.5), an addressee would have to be familiar with the context in which (1.5) was uttered. However, it seems more probable that the key to answering the question, 'too grey for what?' lies in recovering ellipsed material, rather than in resolving any vagueness in *too*. A fully propositional form of (1.5) would have to make explicit any ellipsed material:

(1.6) The bat discovered by Peter is too grey to be a fruit bat.

Despite the distinctions drawn between reference assignment, disambiguation, resolution of vagueness and the recovery of ellipsed material, these all contribute to the same inferential (pragmatic) operation, that of relating a linguistically decoded semantic representation (logical form) to a context. In view of this, and because this study will primarily be concerned with phase two of utterance interpretation (disambiguation, although inferential, I have treated as operating at phase one), I shall use *inferential enrichment* as a blanket term for the pragmatic task of relating logical form and context (as Sperber & Wilson also do at times, as for example in the opening quotation from S&W (1986:174)). I also prefer to follow Recanati (1989:299) in using the terms *semantically underdetermined* and *semantic underdetermination* in place of Sperber & Wilson's *semantically incomplete* and *semantic incompleteness*. This reflects both the unified treatment of reference assignment and inferential enrichment proper in this study, and the characterisation of logical forms as complete semantic representations (albeit inadequate for communicative purposes in isolation from an appropriate context).

The consequences of adopting inferential enrichment are perhaps best illustrated in comparison with an alternative model. For this purpose I shall compare relevance theory with Grice's model, this being a well-known pragmatic theory, and one with which the early development of relevance theory was closely connected.

### 1.3.3 Gricean pragmatics and relevance theory

Grice's observation that communication is possible even in the absence of a code is at the heart of Sperber & Wilson's fundamental claim for the primacy of ostensive-inferential over coded communication (see §1.1). S&W (1986:37-8) makes clear that although Grice's thesis is central, the subsequent development of relevance theory is distinct from that of traditional Gricean pragmatics. There are a number of differences between relevance theory and Gricean pragmatics, the most important being that whereas Grice posits a distinction between what is (literally) said (governed by a code model) and what is implicated (governed by the Co-operative Principle and maxims, and also by decoding in the case of 'conventional implicatures'), relevance theory assumes both implicit and explicit content (in Gricean terminology) to be inferentially determined, with the result that:

“the distinction between semantics and pragmatics - that is, the distinction between linguistically determined meaning and meaning that is contextually determined in accordance with general principles of communication - is not coextensive with the distinction between what is said and what is implicated.”

(Blakemore 1987:34)

Sperber & Wilson's reservations about Grice's explicit/implicit distinction date back at least to 1981 when they suggested that the scope of pragmatics is wider than Grice thought (S&W 1981:159). For Grice, all pragmatic aspects of interpretation, bar disambiguation and reference assignment, are necessarily implicatures, but according to relevance theory, pragmatic interpretation in the form of inferential enrichment of logical form is often required in order to recover explicitly communicated information, or explicatures, in addition to implicatures. To use Gricean terminology, relevance theory posits pragmatic aspects of 'what is literally said'. As Recanati (1989:328) notes, Grice's 'tests' for conversational implicature (cancellability, nondetachability, calculability, and so forth) are not wholly superfluous within the relevance theoretic framework; they still indicate the presence of a pragmatically determined aspect of utterance meaning, but what they do not indicate is whether this is a genuine implicature or a constituent of what is literally said.

This can be illustrated through a consideration of the interpretation of clauses conjoined by *and* (see in particular Carston 1988b, 1993); the examples below (bar (1.10a) and (1.13)) are taken from Wilson & Sperber (1993b). The traditional Gricean system results from conflating the linguistically determined semantic/pragmatic distinction (semantic = linguistically encoded; pragmatic = inferred) and the logically determined semantic/pragmatic distinction (semantic = truth-conditional; pragmatic = non-truth-conditional) and can be summarised in the following two equations (based on Gazdar 1979:2):

- (1.7) a. Truth-conditional = semantic = linguistically encoded  
b. Non-truth-conditional = pragmatic = inferred

Take the examples in (1.8); the temporal and causal connection between the propositions expressed in the conjoined clauses in (1.8) could either be truth-conditional (encoded by *and*) or non-truth-conditional (implied by the maxims of manner and relation):

- (1.8) a. I took out my key and opened the door.  
b. John dropped the glass and it broke.  
c. They planted an acorn and it grew.  
d. Peter left and Mary got angry.

The latter explanation is supported by the examples in (1.9) in which the same temporal and causal interpretations arise in the absence of *and*, and in (1.10) in which there are no such temporal (or, in (1.10a), causal) connections:



- (1.9) a. I took out my key. I opened the door.  
 b. John dropped the glass. It broke.  
 c. They planted an acorn. It grew.  
 d. Peter left. Mary got angry.
- (1.10) a. I like Beethoven and I like Elgar.  
 b. It was dark and I couldn't see.  
 c. Susan is underage and can't drink.

Since the temporal and causal connections of (1.8) are not encoded by *and* they must be pragmatic (and therefore, according to (1.7b), non-truth-conditional) in the Gricean system; this conclusion would be supported by the fact that the temporal and causal interpretations of (1.8d) are cancellable:

- (1.11) Peter left and Mary got angry, but not in that order.

However, sentences such as (1.12a) and (1.12b) indicate that the temporal and causal connections of (1.8d) are in fact truth-conditional; if they were non-truth-conditional (as they must be if pragmatically derived according to (1.7b)), example (1.12a) would be contradictory and example (1.12b) would be tautologous:

- (1.12) a. What happened was not that Peter left and Mary got angry,  
 but that Mary got angry and Peter left.
- b. A: So Peter left and Mary got angry?  
 B: No. Mary got angry and Peter left.

The fact that the above examples are neither contradictory nor tautologous suggests that truth-conditional content need not be linguistically encoded, and cancellability merely indicates that a given interpretation is not linguistically encoded. The equation upon which the Gricean system is based must therefore be abandoned. If, in the divorce settlement between truth-conditionality and linguistic encoding, custody of the term 'semantic' is awarded to linguistic encoding (and therefore the term 'pragmatic' belongs to what is not linguistically encoded), there can be pragmatic aspects of what is literally said.<sup>13</sup>

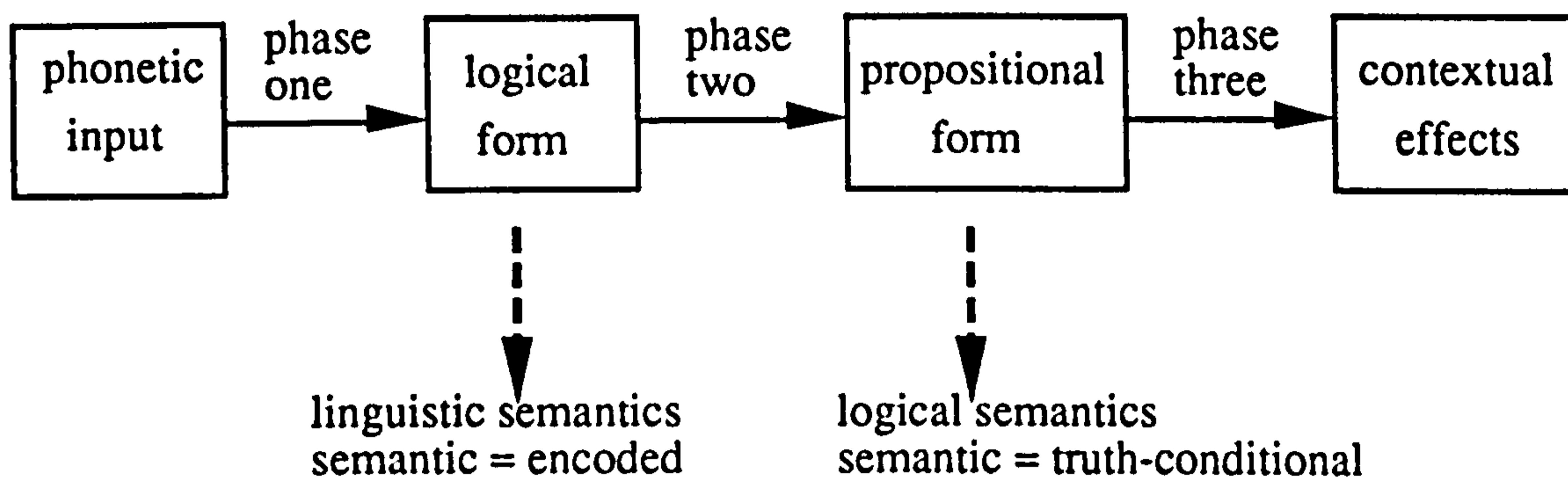
---

<sup>13</sup> The above arguments do not, however, entail that semantically *and* can be equated with logical &. In the preferred reading of (3), B is interpreted as preceding A, whilst in (4) A precedes B:

- |     |    |                  |                             |
|-----|----|------------------|-----------------------------|
|     |    | A                | B                           |
| (3) | a. | The glass broke. | John dropped it.            |
|     | b. | I hit Bill.      | He insulted me.             |
|     | c. | I got caught.    | My best friend betrayed me. |

The relation between linguistic semantics and logical semantics in terms of the three phase model of utterance interpretation which I have been expounding is illustrated in figure 3:

Figure 3: The relation between linguistic and logical semantics



Two further differences between Gricean pragmatics and relevance theory follow from Sperber & Wilson's dissatisfaction with Grice's characterisation of explicit communication, and their own definition of explicitness, given below:

"An assumption communicated by an utterance U is *explicit* if and only if it is a development of a logical form encoded by U."  
(S&W 1986:182)

Firstly, relevance theory assumes *degrees* of explicitness: the greater the element of decoding involved in interpretation, the more explicit the explicature; the greater the degree of inference involved, the less explicit. For example, (1.6) is more explicit than (1.5) (repeated below):

- (1.6)            The bat discovered by Peter is too grey to be a fruit bat.  
 (1.5)            Peter's bat is too grey.

- 
- |     |    |  |   |
|-----|----|--|---|
| (4) | a. | A  | B |
|     | b. | The glass broke and John dropped it.         |   |
|     | c. | I hit Bill and he insulted me.               |   |
|     |    | I got caught and my best friend betrayed me. |   |

In (5) the occurrence of *then* cancels the causal connection between A and B:

- |     |    |   |   |
|-----|----|---|---|
| (5) | a. | A   | B |
|     | b. | I spoke to John and discovered that he was charming.      |   |
|     |    | I spoke to John and then discovered that he was charming. |   |

Taking the above examples along with (1.8) and (1.10) I propose the following description of how *and* functions: If there is a temporal connection between the propositions expressed by two clauses conjoined by *and* then there is also a causal connection, unless e.g. *then* is present; if there is a causal connection there need not be a temporal one.

Secondly, Grice says very little concerning how propositional attitudes, that is, non-truth-conditional functions such as illocutionary force, are communicated. Grice is unclear regarding what he regards as explicit or implicit here, but in relevance theory, these propositional attitudes, also termed 'higher-level explicatures', are incorporated into the same system as other explicatures. In common with speech-act theory and functional grammar, it is recognised that semantic differences between declarative sentences and their non-declarative counterparts are not differences in the propositions expressed, but in the higher-level explicatures communicated, which specify the intended illocutionary-force of the utterance by indicating the direction in which relevance is to be sought (S&W 1986:254). Relevance, and its role in communication (in particular the criterion of consistency with the principle of relevance, which an interpretation must meet in order to achieve contextual effects), is discussed in §1.4 below.

## **1.4 Phase Three: Relevance**

### **1.4.0 Introduction**

I noted in the introduction and in §1.1 that relevance theory is concerned with the interpretation of utterances within a context, cognitively determined, and that the aim of communication is to achieve contextual effects, that is, alterations to the addressee's cognitive environment. This occurs at phase three of utterance interpretation; procedural encoding was first posited as functioning at this phase of the utterance interpretation process, where it constrains the computation of contextual effects, thereby reducing processing effort. In this final section of chapter 1, I shall begin (§1.4.1) by discussing contextual effects and processing effort, before considering their role in determining relevance (§1.4.2). Having clarified the notion of relevance, I shall consider (in §1.4.3) the implications of the central tenet of relevance theory: the principle of relevance, for a theory of ostensive-inferential communication, before summarising the discussion (§1.4.4).

### **1.4.1 Contextual effects and processing effort**

The end result of utterance interpretation is what Sperber & Wilson call a contextual effect. When a set of assumptions is made mutually manifest (see §1.1) by an act of communication it can, when combined with a context, give rise to a further assumption or assumptions. S&W (1986:108) expresses this in the following terms:

"a deduction based on the union of new information {P} and old information {C} is a *contextualisation* of {P} in {C}. Such a contextualisation may give rise to what we will call *contextual effects*."

Contextual effects come in three general forms: implications, strengthening and contradiction. Contextual implications are propositions which are implied neither by {P} alone nor by {C} alone but by the union of {P} and {C}. Strengthening occurs when the union of {P} and {C} increases the probability, that is the strength, of some proposition or propositions contained within {C}, and contradiction occurs when the union of {P} and {C} eliminates a proposition or propositions within {C}.

No matter how contextual effects are derived, a certain amount of processing effort is involved in computing them. Processing effort occurs both at phase one of utterance interpretation in the form of the effort required to decode linguistic input, and at the inferential phases of utterance interpretation. Various factors interact to influence both sources of processing effort. In decoding, factors such as frequency of mention of a lexical item affect its accessibility and hence the processing effort needed to process it, but because linguistic decoding is a modular phase it incurs little processing effort in comparison with inferential processing, as evidenced by the speed with which linguistic decoding occurs (around 250ms in close shadowing tasks). At the inferential phases, the construction of a context - that is, the selection of a subset of old information {C} with which to combine new information {P} - and the determination of the precise relation between {C} and {P}, both incur processing effort. (In chapter two an explicit link will be made between procedural encoding and the reduction of processing effort.)

#### 1.4.2 The principle of relevance

The relation between contextual effects and the processing effort required to derive them forms the basis of the *principle of relevance* to be described below. 'Relevance', as Sperber & Wilson use the term, describes a clearly defined theoretical concept. In this, it is not to be identified with Grice's vague and undefined (Grice 1989:27) use of the word in the Maxim of Relation: "be relevant". S&W (1986) provides three partial and complementary definitions of relevance: relevance in a context, relevance to an individual, and relevance of a phenomenon. I shall look at each in turn.

##### *Relevance in a context*

An assumption is relevant in a context iff it has some contextual effect in that context (S&W 1986:122), and the greater the contextual effects, the greater the relevance. However, computing the contextual effects of an utterance requires some mental effort; it is assumed that human beings aim to process information as efficiently as possible, that is, to minimise the mental effort required, therefore the greater the mental effort required to compute the contextual effects of an utterance, the smaller

the relevance of the assumption identified in the particular context. Contexts are sets of mutually manifest assumptions (§1.1), and as such are dynamic in that they develop as assumptions are made mutually manifest. Since a large number of assumptions are typically mutually manifest in a discourse, and the inferential phase of utterance interpretation, being global, has potential access to all of these, some criterion for the selection of the particular context within which a given utterance is to be processed is required (similar to the inhibitory effect of inferential processing on linguistic decoding, mentioned in §1.2). The assumption that speakers aim to be relevant is such a criterion, and this remains constant; according to relevance theory contexts vary but speakers always aim to be relevant (S&W 1986:142).

This entails that a sentence might be more relevant if uttered in one context than in another. For example, (1.5) uttered in the context of a zoological field trip would activate a number of mutually manifest assumptions (for example that Peter had been looking for fruit bats, and that fruit bats are brown) and so achieve the intended contextual effect without incurring disproportionate processing effort.

(1.5) Peter's bat is too grey.

Uttered in a context where such assumptions were not mutually manifest, (1.5) would not be as relevant as (1.6) would:

(1.6) The bat discovered by Peter is too grey to be a fruit bat.

However, in a context where all the appropriate assumptions were mutually manifest an utterance of (1.6) would cause the addressee excess processing effort, in particular at phase one (linguistic decoding) and would not be so relevant.

#### *Relevance to an individual*

The First (Cognitive) Principle of Relevance (S&W 1995) states that human cognition tends to be geared towards the maximisation of relevance:

“Other things being equal, the greater the cognitive effect achieved by the processing of a given piece of information, the greater its relevance for the individual who processes it [...and...] the greater the effort involved in the processing of a given piece of information, the smaller its relevance for the individual who processes it.”

(Wilson & Sperber 1988:140)

A ‘cognitive effect’ is a contextual effect in an individual, and should be taken to include social as well as purely propositional relevance, that is, relevance to an individual’s particular goals and purposes. Sanders (1988:607) criticised relevance

theory for only taking account of relevance to the individual's prior cognitions, at the expense of relevance to the social activity in progress. Similar criticisms are made by Mey & Talbot (1988), Nemo (1988) and O'Neill (1988). However, Sperber & Wilson themselves are aware of the affective dimension in communication and explain (Sperber & Wilson 1987:742) that the social dimension of relevance theory is poorly articulated simply because they had no idea of how to define relevance to a purpose, goal, etc. in social terms.

Also, the degree of relevance achieved is dependent on the addressee as well as the speaker. For example, a lecturer may aim to be optimally relevant with regard to her audience, that is, for the content of her lecture to bring about substantial contextual effects whilst being as perspicuous as possible; however, some students will be unwilling to make the necessary processing effort, and others may already be familiar with the content, so degrees of relevance achieved will vary within the audience. The link between individuals and phenomena is relevance; phenomena which are unlikely to be relevant tend to be filtered out by human perceptual mechanisms; that is, the perceptual mechanisms are relevance-oriented (S&W 1986:152). This leads us to the final partial definition of relevance:

#### *Relevance of a phenomenon*

"A phenomenon is relevant to an individual to the extent that the contextual effects<sup>14</sup> achieved when it is optimally processed are large [...and...] the effort required to process it optimally is small."  
(S&W 1986:153)

The only phenomena which individuals assume always have some relevance are *ostensive stimuli*. Ostensive stimuli attract the audience's attention and focus it on the communicator's intentions (S&W 1986:153-4). They include not only speech but other deliberate gestures which also come with a guarantee of relevance.

#### *The principle of relevance*

The above definitions of relevance combine to give the following informal definition of *optimal relevance*:

An utterance, on a given interpretation, is optimally relevant if and only if:

- (a) it achieves enough effects to be worth the hearer's attention;
- (b) it puts the hearer to no gratuitous effort in achieving those effects.

(Smith & Wilson 1992:5)

---

<sup>14</sup> Revised (S&W 1995) to read 'positive cognitive effects', where 'positive' reflects the cognitive goal of deriving only true assumptions from input.

Combining the presumption of relevance guaranteed by ostensive stimuli with the above definition of optimal relevance results in the *presumption of optimal relevance* which, as Carston (1988a:60) points out, is a presumption of *adequate* (rather than maximum) cognitive effects for least processing effort. This is better known as the *principle of relevance*<sup>15</sup>:

“Every act of ostensive communication communicates the presumption of its own optimal relevance.”

(S&W 1986:158)

Before looking at the implications of the principle of relevance for a model of utterance interpretation, I shall briefly mention a few differences between it and traditional Gricean assumptions. Whilst the principle of relevance is an exceptionless generalisation about communication which cannot be overtly violated (S&W 1986:162, Wilson & Sperber 1988:140), Grice's Co-operative Principle and maxims are conversational norms which can be violated;<sup>16</sup> the nature and purpose of these violations is culture specific, as Keenan's (1976) study of violations of the maxim of quantity by some Malagasy interlocutors demonstrates. Also, Clark (1982:126) noted that an early formulation of the principle of relevance “explicitly allows both authorized and unauthorized inferences... therefore, it cannot be used as a specification of Grice's maxim of relation. It does not belong to Grice's co-operative principle”, a claim with which Sperber & Wilson (1982:131) agreed. Although the principle of relevance has subsequently been criticised for being vague (e.g. Mey & Talbot 1988, Gorayska & Lindsay 1993), the maxim of relation remains undefined (Grice 1989:27).

### 1.4.3 Consistency with the principle of relevance

The central consequence of applying the principle of relevance is that the only interpretation whose selection it warrants is the first one tested and found to be consistent with the principle. This is not as gratuitous as it at first appears, but is rather a result of the way information is processed. Broadly speaking, there are two

---

<sup>15</sup> This is strictly speaking the Second (Communicative) Principle of Relevance and is grounded in the First (Cognitive) Principle of Relevance. As in S&W (1995) the phrase ‘principle of relevance’ refers to the second (communicative) principle. S&W (1995) also modifies the presumption of optimal relevance, acknowledging “the speaker's right to be lazy or prudish, i.e. to have her own preferences and take them into account.” The definitions given here are adequate for the purposes of this thesis.

<sup>16</sup> This distinction has developed: “Earlier versions of relevance theory were closer in these respects to Grice's approach. We had generally assumed that there was a presumption of maximal rather than optimal relevance, and that communicator and audience had to have and use knowledge of the principle of relevance. However, the idea that the principle was exceptionless was there from the start” (S&W 1986:261 n.11). Maximal rather than optimal relevance is described in Sperber & Wilson (1982, 1985/6), and is referred to in e.g. Wilson & Sperber (1988:140), Jucker (1993).

possible strategies for information processing: listing and ranking hypotheses, and testing hypotheses one by one. The former is psychologically unfeasible, as Sperber & Wilson note:

“If the only way of finding the right interpretation were to list and rank all possible interpretations, then all possible interpretations would require the same amount of effort: namely, the effort needed to construct and compare them. It is hard to think of any ostensive stimulus that would be worth such an absurd amount of effort.”

(S&W 1986:166)

The second strategy also appears problematic, since there are various interpretations which might be consistent with the principle of relevance. If we simply select the first appropriate hypothesis, this seems to beg the question, since there must be some method for choosing in which order to test hypotheses. Sperber & Wilson avoid this dilemma by appealing to the fact that *contexts* are variable (whereas speakers consistently aim at optimal relevance), and it is not simply the ostensive stimulus but the union of stimulus and context which ultimately gives rise to contextual effects. Ostensive stimuli give access to "assumption schemas", that is, logical forms stored systematically in the memory, and these are then combined with other aspects of the context to form full hypotheses. The most accessible context at a given point in discourse consists of the most salient mutually manifest assumptions, determined as assumptions derived from previously interpreted discourse and the current physical environment as well as other prior assumptions and beliefs.

Given this combination of context and stimulus, some hypotheses are more accessible than others, and therefore require less processing effort (S&W 1986:167). The most accessible interpretation, since it requires least processing effort and is therefore processed quickest, is formulated and tested first; if it achieves adequate contextual effects it is consistent with the principle of relevance and the interpretation process ceases. Therefore the only interpretation warranted by the principle of relevance is the first one tested and found consistent with it.

As noted above, an utterance may fail to be optimally relevant, but speakers will always aim for optimal relevance. Combined with the first-past-the-post model of utterance interpretation, this is the basis of the pragmatic criterion of consistency with the principle of relevance:

“for an utterance to be understood, it must have one and only one interpretation consistent with the fact that the speaker intended it to seem relevant to the hearer - adequately relevant on the effect side and maximally relevant on the effort side. We will say that in this case the interpretation is *consistent with the principle of relevance...*”

(Wilson & Sperber 1988:141)



The criterion of consistency with the principle of relevance provides the answer to the question posed in §1.1, of how it is that communication is so often successful, and also resolves the frame problem (Pylyshyn 1987), namely that if the inferential system automatically considered all potentially relevant data "this would trigger a computational explosion with no end in sight" (Sperber 1994:43). Since speakers are assumed to be aiming to make their utterances consistent with the principle of relevance, audiences can be fairly confident that the first interpretation tested and found to be consistent with the principle is the one intended by the speaker. There is thus no need to reserve judgement on this interpretation or to go on infinitely formulating and testing alternative interpretations since "the interpretation process has an inbuilt stopping place" (Smith & Wilson 1992:7).

#### *Relevance-driven interpretation: the case of metaphor*

The criterion of consistency with the principle of relevance underlies the claim made in §1.1.2 above that utterance interpretation is essentially relevance-driven. It also entails that the three phases of utterance interpretation need not occur sequentially in the order presented here (i.e. first phase one, followed by phase two, and finally phase three). Evidence cited earlier (Crain & Steedman 1985, Altmann & Steedman 1988, Marslen-Wilson & Tyler 1987, Marslen-Wilson *et al* 1993) suggests that there might be no fixed order for the operation of the three phases of utterance interpretation. This principle is exploited in the relevance theoretic account of metaphor and is supported by experimental data (Gibbs 1986a, b, 1989). In order to illustrate how this works in practice I shall briefly discuss how metaphor can be treated within a relevance theoretic framework.

A model of metaphor must explain how metaphors are recognised as such and how they are processed. Stroik (1988) gives the following account of how metaphor is treated in semantic and traditional pragmatic models. In a decompositional semantic model, the metaphor "All men are pigs" could be analysed by substituting one or more of the constituent properties of the word *pig* into the definition of *man*. However, such an account neither specifies which constituent property to substitute nor accounts for non-truth-conditional connotations (e.g. that pigs are usually viewed as fat and dirty) and hence cannot cope with literally self-contradictory metaphors such as "Pigs are not necessarily pigs".

Of traditional pragmatic models, Stroik (1988) considers speech act theories and Gricean pragmatics. According to speech act theories, when an utterance is known to be literally false, it may be reinterpreted as if it were another, true, utterance on the

assumption that the speaker intends to be meaningful. However, Stroik (1988:21) argues that such accounts “neither capture the conditions under which metaphors are identified nor formulate a procedure for expressing the extra-literal meaning necessarily associated with metaphor.” That is, speech act theories fail to account for tautologies (“Boys will be boys”), predict that all obviously false utterances (“The square root of forty nine is six” etc.) can be metaphors, and cannot assign interpretations to metaphors, hence Stroik favours a Gricean account. According to this model, metaphors may “flout”, that is, “blatantly fail to fulfil” (Grice 1975:49), the maxims of quantity (“Boys will be boys”: uninformative) or quality (“All men are pigs”: semantically false) without violating the CP. Metaphors engender conversational (weak) implicatures, which are cancellable (“Bill isn’t fat but he’s still a pig”), thus allowing the hearer to assume that a maxim is being “exploited”.

The Gricean account of metaphor interpretation tacitly assumes that the three phases of utterance interpretation posited here occur in strict sequence and hence involves the following stages: the utterance is given a literal interpretation, this is seen to violate the CP, the flouted maxim is identified, various weak implicatures are generated, these are disposed amongst with reference to the CP. All this would suggest that metaphors should be harder (and hence take longer) to process than non-metaphorical utterances, but experimental evidence (Gibbs 1986a, b, 1989) suggests that this is not the case. If the interpretation process is relevance-driven, on the other hand, the initial interpretation will not result purely from the literal interpretation of the utterance, but will be restricted by the utterance interpretation context and the principle of relevance. Hence, processing time is not wasted deriving and rejecting literal interpretations and formulating alternatives, that is:

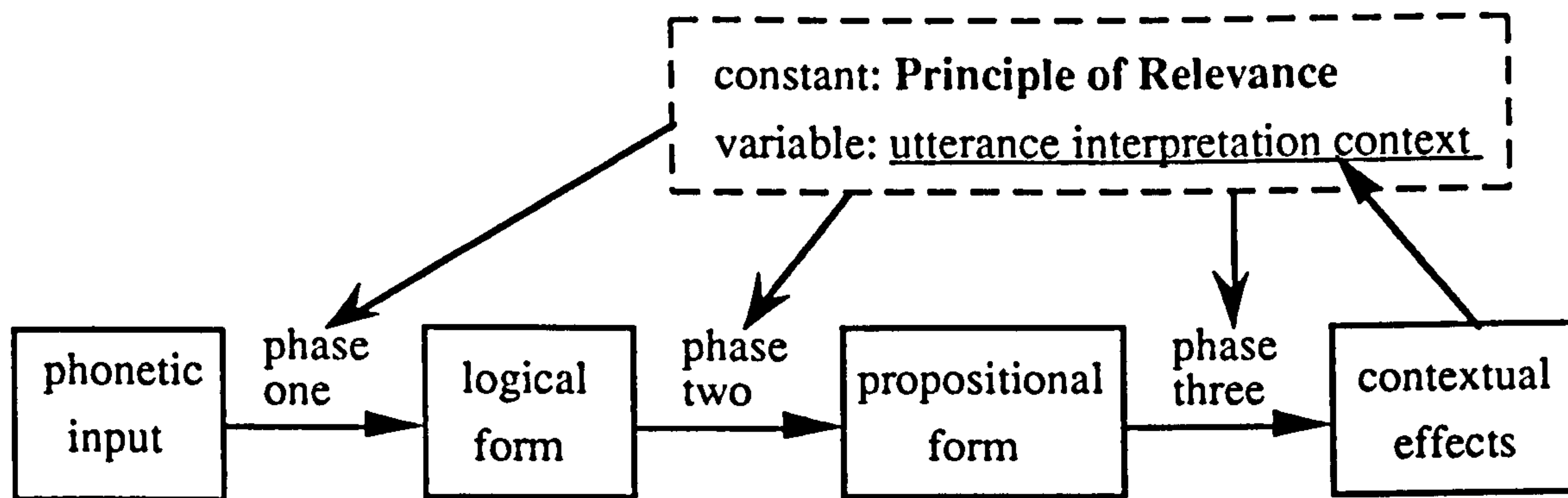
“By the criterion of consistency with the principle of relevance, an utterance will be understood as loose or metaphorical unless nothing less than a fully literal interpretation will do.”

(Wilson & Sperber 1988:144)

#### 1.4.4 Summary

The simple model of the three phases of utterance interpretation presented in figure 1 (§1.1.2) can now be augmented to take account of the relevance-driven nature of utterance interpretation as described above. Figure 4, below, is identical to figure 1 apart from the additional representation of the principle of relevance and the utterance interpretation context.

Figure 4: The utterance interpretation process



The arrow from the combined principle of relevance/utterance interpretation context box to phase one of utterance interpretation represents the process occurring during interpretation of, for example, metaphor; this inhibits the interpretation process from deriving a literal (linguistically decoded) interpretation of an utterance if a loose or metaphorical interpretation is more relevant. The arrow to phase two facilitates correct inferential enrichment of underdetermined aspects of a logical form, for example in cases of reference assignment (§1.3.2). The arrow to phase three of utterance interpretation helps determine the intended contextual effects which in turn feed back into the utterance interpretation context, so that the whole system is being constantly updated.

## 2 CONCEPTUAL AND PROCEDURAL ENCODING

In chapter 1 I outlined the relevance theoretic model of utterance interpretation and characterised it in terms of three phases, phases two and three of which are purely inferential. In this chapter I turn to the central topic of this thesis: the distinction between conceptual and procedural encoding. §2.1 explains what is meant by these types of encoding and provides arguments, from within the relevance theoretic framework outlined above, both for and against such a distinction, before briefly discussing the different functions which the conceptual/procedural distinction might serve at phases two and three. §2.2 proceeds to exemplify procedural encoding at phase three (constraining the inferences to be drawn from propositional conceptual representations) drawing on work already conducted in this field. Finally, in §2.3, I propose a definition of procedural encoding at phase three along with criteria by which procedural and conceptual encoding at phase three can be distinguished.

### 2.1 Information Types in Linguistic Semantics

#### 2.1.0 Introduction

It was noted in §1.3.3 that linguistic semantics, in which semantic information is defined as that which is encoded and pragmatic information as that which is inferred, is not coextensive with logical semantics, in which semantics is concerned with what is truth-conditional and pragmatic with what is non-truth-conditional. As a result, the distinction between truth-conditional and non-truth-conditional information is of only marginal importance for linguistic semantics. Of greater significance are the two basic distinctions recognised by relevance theory: First, there is the distinction discussed in §1.3.3 between explicit information, involving an element of encoding - since "An assumption communicated by an utterance U is *explicit* if and only if it is a development of a logical form encoded by U" (S&W 1986:182) - and implicit information which is purely inferential.

Secondly, there is the distinction between conceptual encoding and procedural encoding. Conceptual information contributes to contextual effects by specifying the content of an utterance at the level of logical form, which is enriched to give rise to new information ({P} - which is contextualised in {C}: old information). Procedural information does not contribute directly to the propositional content of an utterance, rather its contribution to the interpretation process lies in reducing processing effort; at phase two of utterance interpretation procedural information constrains the processes of inferential enrichment (in ways to be discussed in chapter 3), whilst at

phase three it aids the addressee's search for contextual effects. Before discussing how procedural information functions at phase three, I shall look in greater detail at the arguments for and against the conceptual/procedural distinction.

### 2.1.1 Arguments for a dualistic semantics

The role of procedural encoding at phase three can be compared to that of a catalyst in a chemical reaction, in that it facilitates the computation of the intended contextual effects without directly contributing to them, in much the same way that a catalyst induces or accelerates a chemical reaction without itself being part of that reaction.

As discussed in §1.1.1 and 1.4.3, relevance theory is largely concerned with the context dependence of utterance interpretation (and derivatively of text interpretation). An utterance interpretation context is viewed as a subset of the addressee's cognitive environment, consisting of the most salient mutually manifest assumptions. Contexts are therefore dynamic; an addressee's cognitive environment is constantly updated as new assumptions are made, and different assumptions become more or less salient. The particular subset of the addressee's cognitive environment within which utterance interpretation occurs is therefore unstable. In contrast to the instability of contexts and the dynamic nature of context formation, the presumption of optimal relevance provides a constant reference point for the interpretation process. Much formal semantics (e.g. Lewis 1972), on the other hand, defines context "objectively, independently of speaker or hearer, as the set of individuals, times, places, etc. that supply the values of the variable expressions" (Blakemore 1987:9).<sup>17</sup>

The characterisation of utterance interpretation provided in chapter 1 suggests that during the inferential phases of utterance interpretation an addressee performs the following three tasks: 1) the construction of an appropriate interpretation context, 2) the correct relation of logical form to context in order to arrive at the intended propositional forms, and 3) the correct relation of propositions to context in order to derive adequate contextual effects. Of these, the third task takes precedence over the other two; if the context initially selected as the utterance interpretation context fails to yield adequate contextual effects, an alternative interpretation context must be sought which does; or if the propositional form derived via inferential enrichment fails to combine with existing assumptions to yield adequate contextual effects, an alternative interpretation may be sought. This follows from the characterisation (in

---

<sup>17</sup> An exception is Situation Theory which, according to Kempson (1994b:143), is, apart from relevance theory, the only theory to propose a general solution to the problem of context dependence. (See Kempson 1994b for discussion.)

§1.4.3) of utterance interpretation as being relevance-driven: every act of ostensive communication communicates the presumption of its own optimal relevance. But, as we have seen, there are two sides to optimal relevance, since optimal relevance is defined in terms of adequate contextual effects and minimal processing effort (§1.4.1). It is to be expected, therefore, that speakers will employ strategies aimed at reducing addressees' processing requirements, by indicating in what way propositional forms are intended to achieve contextual effects.

This is the basis of Blakemore's (1987) argument that some expressions, including *so*, *too*, *after all* and *because*, "impose constraints on the context in which the utterances containing them must be interpreted" (1987:75). A dualistic semantics, in which Gricean conventional implicatures are viewed as imposing constraints on the contexts in which utterances containing them may occur, was proposed by Stalnaker (1974:212). However, as S&W (1986:263 n.21) observed, in the absence of an account of the role of context in utterance interpretation, it was difficult to see why such structures should exist. Blakemore (1987, 1988a, 1990) noted that since, in a relevance theoretic framework, the impact of an utterance depends on combining the proposition presented with a context (1988a:187) as effectively as possible (i.e. for a minimal amount of processing effort), the use of structures whose "sole function is to guide the interpretation process by specifying certain properties of context and contextual effects" (1987:77) is to be expected. She concludes:

"This suggests a non-unitary theory of linguistic semantics. On the one hand, there is the essentially *conceptual* theory that deals with the way in which elements of linguistic structure map onto concepts - that is, onto constituents of propositional representations that undergo computations. On the other, there is the essentially *procedural* theory that deals with the way in which elements of linguistic structure map directly onto computations themselves - that is, onto mental processes."

(Blakemore 1987:144)

The types of computations which *propositional* representations can undergo can be divided into two broad categories: those which result in implicatures (assumptions not directly derived from information represented in logical form) and those which result in higher level explicatures (speech act information). Correspondingly we can expect there to be at least two types of procedural encoding: one which encodes constraints on implicatures and another which encodes constraints on higher-level explicatures. Both types of computation give rise directly to contextual effects, and (since neither contribute to inferential enrichment resulting in the identification of a propositional form, which occurs at phase two) occur at what I have termed phase three of utterance interpretation. Thus we can talk of procedural encoding (functioning) at phase three of utterance interpretation, whose purpose is to constrain the inferences drawn from

propositional representations, and thus aid the search for intended contextual effects. Such procedural encoding will be exemplified in §2.2.

It has further been suggested (Wilson & Sperber 1993a) that procedural encoding may function at the level of linguistic semantic representation (that is, at phase two of utterance interpretation) to constrain the *construction* of propositional representations (the expression of experiential or ideational meaning in the terminology of functional grammar). I shall begin this discussion of Wilson & Sperber's proposals by clarifying some terminological vagueness, and in the process bring their account into line with my three-phase model of utterance interpretation. I shall then consider their extension of procedural encoding to include truth-conditional expressions.

According to Wilson & Sperber (1993a:2) the distinction between conceptual and procedural information corresponds to the distinction between the construction and the manipulation of conceptual representations:

"inferential comprehension involves the construction and manipulation of conceptual representations. An utterance can thus be expected to encode two basic types of information: representational and computational, or conceptual and procedural - that is, information about the representations to be manipulated, and information about how to manipulate them."

(Wilson & Sperber 1993a:2)

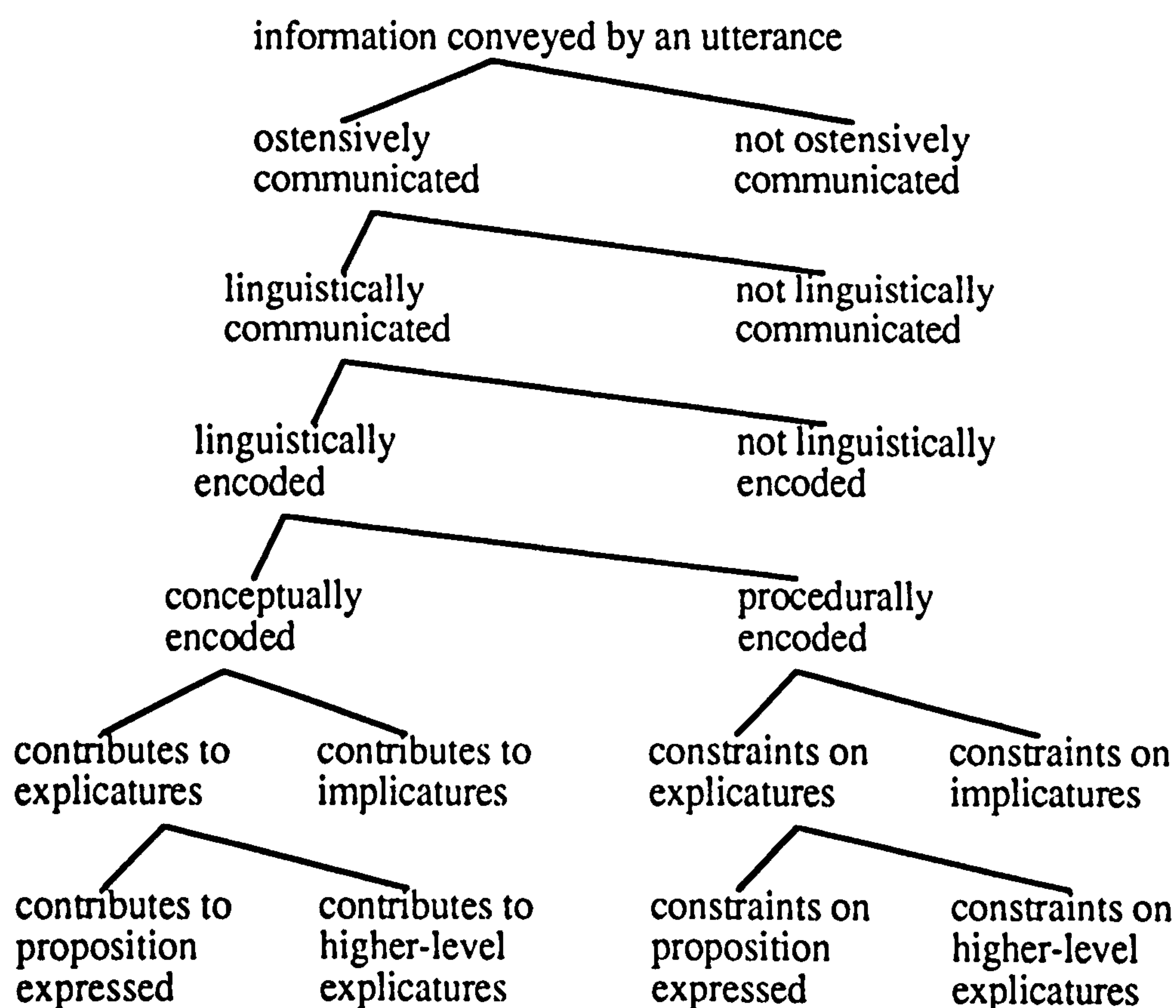
Wilson & Sperber (1993a) use the term 'conceptual representation' to refer to two distinct types of mental object - one a representation of a proposition and the other a sub-propositional representation. This can be demonstrated as follows: Wilson & Sperber (1993a:2) state that "some truth-conditional constructions encode concepts, others encode procedures," where a construction is truth-conditional "if and only if it contributes to the proposition expressed" (W&S 1993a:6). Thus, a conceptual representation whose manipulation is constrained by a truth-conditional construction must be less than fully propositional, since its manipulation contributes to the proposition of which it is part. Yet procedural constructions functioning at phase three (which Wilson & Sperber term procedural and non-truth-conditional) constrain the manipulation of fully propositional conceptual representations. Wilson & Sperber (1993a) thus fail to make explicit that the term 'conceptual representation' refers to both propositional and sub-propositional conceptual representations.<sup>18</sup>

---

<sup>18</sup> This lack of clarity is grounded in Wilson & Sperber's conflation of what I have termed phase two and phase three of utterance interpretation: "Our book *Relevance* (Sperber and Wilson 1986) treats utterance interpretation as a two-phase process: a modular decoding phase is seen as providing input to a central inferential phase in which a linguistically encoded logical form is contextually enriched and used to construct a hypothesis about the speaker's informative intention." (Wilson & Sperber 1993a:1)

Wilson & Sperber note that the conceptual/procedural distinction is not coextensive with the traditional truth-conditional vs. non-truth-conditional distinction, nor with the other relevance theoretic distinction between explicit and implicit information. Explicit information concerns not only propositions, which have truth conditions, but also information about the speech acts an utterance is intended to perform, which is illocutionary as opposed to propositional, and therefore non-truth-conditional. Given that the conceptual/procedural distinction cross-cuts both the other distinctions, Sperber & Wilson (1993a) recognise six subcategories of linguistically encoded information, corresponding to the six 'head' categories in figure 5 below:

Figure 5: Types of communicated information.



(Wilson & Sperber 1993a:3)

Below, I shall briefly discuss the six subcategories of linguistically encoded information posited by Wilson & Sperber (1993a), addressing first linguistically encoded conceptual information and then linguistically encoded procedural information.

Conceptually encoded information which contributes to explicatures is viewed as being of two kinds: truth-conditional, which contributes to the proposition expressed, and non-truth-conditional, which contributes to higher-level explicatures. The former includes "most regular 'content' words" (W&S 1993a:19); the latter includes, for



example, illocutionary adverbials such as *seriously*, *frankly* and *confidentially*, which are not constituents of the proposition expressed (W&S 1993a:16-19).

One further logically possible subcategory of linguistically encoded information which is included in figure 5 but not discussed elsewhere in W&S (1993a) is conceptually encoded information which contributes to implicatures. Deirdre Wilson (p.c. 16/6/94) has stated that, although this subcategory was left open, perhaps nothing fits in it. She has suggested that one might want to claim that, in Gricean terms, (2.1) conventionally implicates (2.2):

(2.1) John isn't here yet.

(2.2) John is expected.

or that (2.3) conventionally implicates (2.4),

(2.3) I didn't manage to finish the paper.

(2.4) I tried to finish the paper.

If this is so, words like *yet* and *manage* could be characterised as carrying conceptually encoded implicatures (either implicated premises or implicated conclusions). However, to claim that there are words that regularly carry a class of conceptually encoded implications is contradictory; decoding provides the 'skeleton' of what is explicitly communicated, so unless a clear distinction can be drawn between "regularly carrying" and "encoding" no clear distinction can be made between implicatures and explicatures. In relevance theory, "explicit" is a degree term: the greater the amount of decoding involved in interpretation, the more explicit an explicature; the greater the degree of inference involved, the less explicit. This suggests that (2.2) and (2.4) should be treated as explicatures derived from utterances of (2.1) and (2.3) respectively, through a combination of inference and decoding (of *yet* and *manage* respectively). Even if the degree of decoding required were minimal in comparison to the amount of inference, these would still be cases of explicit communication. I conclude, therefore, that the subcategory of conceptually encoded information which contributes to implicatures is logically impossible.

We turn now to linguistically encoded procedural information. Procedurally encoded information which imposes constraints on implicatures is non-truth-conditional and includes discourse connectives such as *so* and *well* (to be discussed below) which guide for intended contexts and contextual effects. Procedurally encoded information which imposes constraints on explicatures can also be non-truth-conditional, for

example illocutionary force indicators such as non-declarative syntactic mood, and the Kiswahili interrogative particle JE (see §2.2.3) which constrain higher-level explicatures (see §1.3.3). Both kinds of non-truth-conditional procedurally encoded information were characterised above as procedural encoding functioning at phase three of utterance interpretation. In addition, Wilson & Sperber (1993a:20) suggest that there is a subcategory of linguistically encoded information which imposes constraints on explicatures and is truth-conditional, the purpose of which is to constrain the proposition expressed. They propose that procedural encoding of this sort (which I have termed procedural encoding at phase two) includes *I* and other personal pronouns, which "guide the search for the intended referent, which is part of the proposition expressed" (W&S 1993a:21).

Implicit in Wilson & Sperber (1993a) is the suggestion that, since there are six logically possible subcategories of linguistically encoded information, we can expect there to be exponents of each category. I have just shown, however, that the subcategory of conceptually encoded information which contributes to implicatures is not, in fact, logically possible and therefore nothing can be expected to be an exponent of it. The other types of conceptual encoding are well attested, as are the two types of procedural encoding at phase three (see §2.2). Now that the symmetry has been broken, we would be justified in exhibiting scepticism with regard to the existence of the remaining subcategory - procedural encoding at phase two - exponents of which are not nearly so well attested. Further discussion of procedural encoding at phase two - its role in the construction of fully propositional conceptual representations and the type of expressions which might play such a role - will be resumed in chapter 3. Before then, I shall address some initial objections to the notion of procedural encoding in general (§2.1.2) and look at examples of procedural encoding at phase three (§2.2).

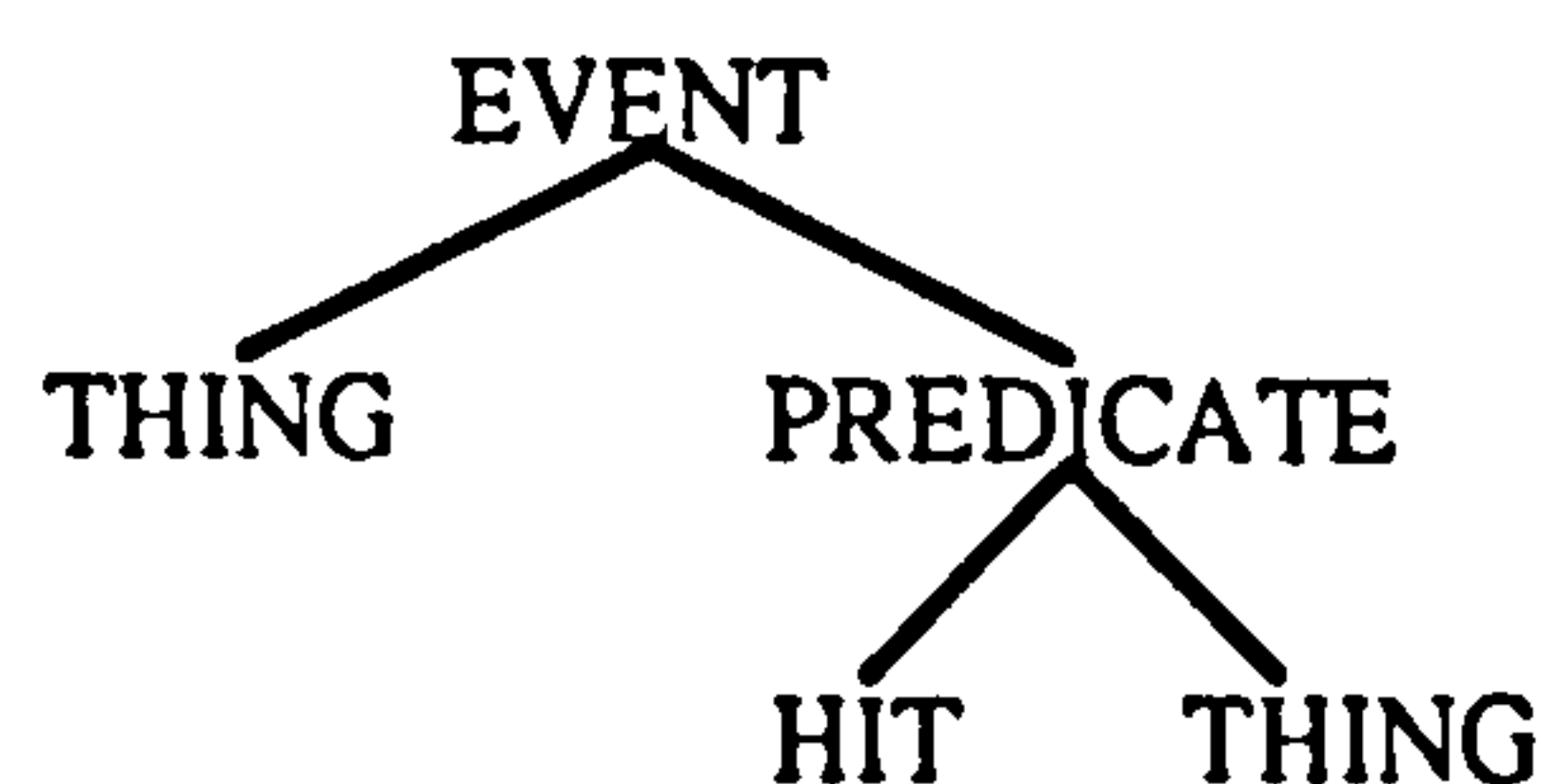
### 2.1.2 A reappraisal of the conceptual/procedural distinction

Before turning to an exemplification of procedural encoding at phase three of utterance interpretation (to constrain the manipulation of propositional conceptual representations), I shall consider an objection to the whole notion of a conceptual/procedural distinction. The arguments with which I shall be concerned are due to Groefsema (1992) and arise from work within the relevance theoretic framework. Groefsema's arguments raise a number of issues which I shall take on board in subsequent investigations into procedural encoding at phase two and the characterisation of different types of procedural information.

Groefsema's arguments against a dualistic semantics assume Jackendoff's (1983) compositional theory of Conceptual Semantics (cf. Jackendoff 1990). Jackendoff proposed that meaning in natural language consists of information structures represented at the level of conceptual structure, the major units of which are conceptual constituents belonging to a small set of ontological categories: THING, EVENT, STATE, ACTION, PLACE, DIRECTION, PATH, PROPERTY, AMOUNT, MANNER (and also possibly: SOUND, SMELL and TIME). For Groefsema, these categories serve to specify the logical categories (variables over conceptual representations) of which S&W (1986) claims logical hypotheses consist.

According to Groefsema (1992:213-220) exponents of what I have termed both conceptual and procedural information are all concepts. Concepts provide access to logical selection frames, which specify: a) what a concept has to combine with to yield a well formed formula in the language of thought, and b) what conceptual category the concept belongs to (if any), and what the concept can combine with (if anything) to form a complex concept. She distinguishes between concepts which occur in conceptual structure (what I have termed conceptual information) and those which do not (what I have termed procedural information), the function of this latter sort of concept being "to constrain the interpretation of the utterance in which they occur in some way." (Groefsema 1992:217)

As an example of a concept which occurs in conceptual structure (and is therefore an example of conceptual information), Groefsema (1992:213-214) takes the verb *hit*, as in *Pat hit Anne*, which gives rise to the concept HIT:



This gives access to the following logical selection frame:

[[EVENT \_ ] [PLACE] [TIME]]  
PROPOSITION

[PREDICATE [HIT [THING]] ([MANNER]) ([PLACE]) ([PROPERTY\*])]

The top tier represents the fact that the concept expresses an event which occurs at a place and a time. The second tier indicates that *hit* is also a predicate which combines with a thing (*Anne*), and optionally with a manner (*violently*), a place (*on the head*), and a property (*with a stick*) - the star after PROPERTY indicates that the resulting

complex predicate can also have a property. Note that the term HIT appears as a constituent of the second tier of this logical selection frame. Concepts which occur in conceptual structure (i.e. which are represented within the logical selection frames to which they give rise) include verbs, nouns and adjectives (Groefsema 1992:215-218).

The complementizer *that* and the determiners *a/an* and *the* also give access to two-tier logical selection frames, but unlike HIT which belongs to the conceptual category PREDICATE, do not belong to any conceptual category - that is, they do not appear within the logical selection frames to which they give access. Groefsema (1992:218) interprets *a(n)* as follows:

what *a(n)* does is signal that an instance of a THING is at stake. In other words, it puts a constraint on the interpretation of a following THING. This means that the phonological or orthographic form *a(n)* maps onto a conceptual address A, which gives access to the following logical selection frame:

[[EVENT/STATE [THING] [PREDICATE] [PLACE] [TIME]]  
PROPOSITION

[THING]  
Individuation:  
Instance

This contrasts with the definite article which gives rise to a logical selection frame identical to that of the indefinite article except that an instance of THING is an accessible<sup>19</sup> instance rather than just an instance:

[[EVENT/STATE [THING] [PREDICATE] [PLACE] [TIME]]  
PROPOSITION

[THING]  
Individuation:  
Accessible instance

Since these expressions do not occur in conceptual structure (that is, they do not appear within the logical selection frames to which they give access), they encode procedural information. This conforms to previous analyses of determiners, such as Hawkins (1978), which treat them as encoding processing instructions (see §3.1 for further discussion). What Groefsema's analysis demonstrates is that both conceptual and procedural information might, in principle, be amenable to semantic representation in terms of the same set of primitive ontological categories represented

---

<sup>19</sup> Groefsema (1992) adopts Kempson's (1988a) model of accessibility in discourse.

as abstract semantic representations in 'the language of thought'.<sup>20</sup> In this, she runs counter to Blakemore (1987) who, in line with S&W (1986), excluded all non-propositional information from logical form. Implicit in Blakemore's theory (according to Groefsema 1992:219) is the suggestion that an expression encoding procedural information "can be an element of linguistic structure, but nevertheless not be part of the linguistic representation that [Blakemore] assumes is computed in the interpretation process [i.e. the logical form]." Groefsema's response is to suggest that both conceptual and procedural information "must be represented in a format that can be 'read' by the central cognitive device, i.e. they must be represented in the language of thought." (1992:220)

What Groefsema assumes is that every element of linguistic structure, once decoded, is represented in the language of thought; this she contrasts with the view that only logical forms, which are "blueprints for propositions" (Blakemore 1987:18), are represented in the language of thought. An alternative view<sup>21</sup> holds that although only conceptual information is represented in the language of thought, this can contribute either to propositions (the information in a logical form does this) or to higher-level explicatures, that is "conceptual representations [which] though true or false in their own right, ... do not generally contribute to the truth conditions of their associated utterances" (W&S 1993a:16). Procedural information, on the other hand, is not represented in the language of thought, but rather consists of constraints which function rather like paths along which the interpretation of conceptual information is channelled. However, it is not obvious how or where procedural information is to be represented according to this view.

I shall follow Groefsema in considering both conceptual and procedural information to be represented in the language of thought, but I shall retain the distinction between conceptual and procedural encoding by viewing the output of phase one of utterance interpretation (linguistic decoding) as giving rise to both conceptual representations and constraints on how those representations are to be manipulated. Crucially, it is whether an expression contains in its logical selection frame information about itself, or about whatever falls within its scope (the manipulation of which it constrains), that determines whether it encodes conceptual or procedural information. This is analogous to the syntactic information encoded by a linguistic expression which may state both the syntactic class of the encoding expression and the syntactic class of any

---

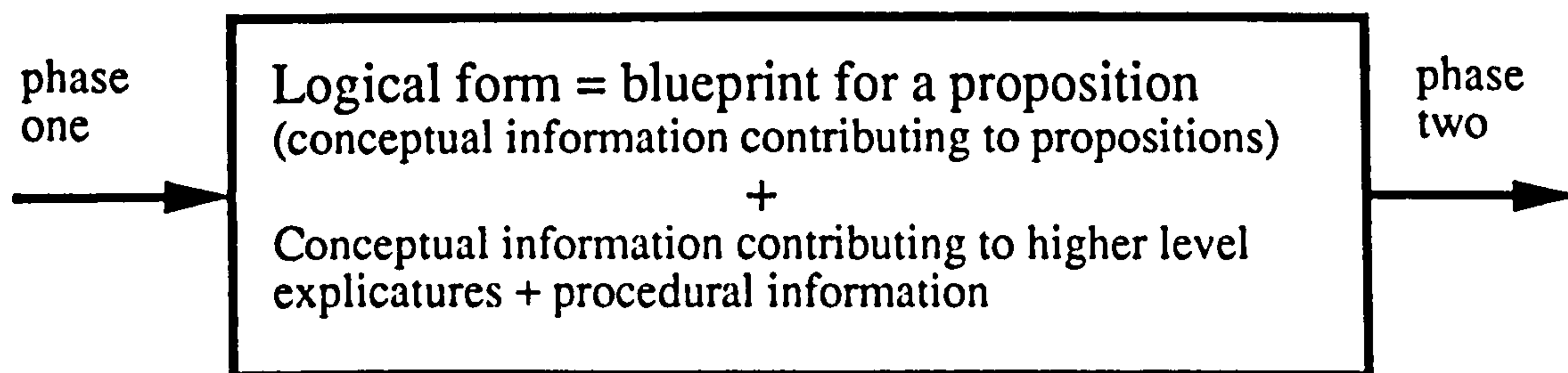
<sup>20</sup> In chapter 5 I propose that, diachronically, exponents of procedural encoding may develop from exponents of conceptual encoding; this type of development could only be possible if conceptual and procedural information were related at some fundamental level of semantic representation.

<sup>21</sup> Deirdre Wilson, p.c. 28/10/95.

element with which the encoding expression may combine to form a well-formed string (subcategorization information).

Thus Groefsema's arguments can be viewed as leading to a reformulation rather than a refutation of the conceptual/procedural distinction. The different types of linguistically encoded information which can be represented in the language of thought (i.e. as the output of phase one of utterance interpretation) are illustrated in figure 6 (figure 1 provides the wider context):

Figure 6: Information types represented in the language of thought



Before continuing with a discussion of procedural encoding at phase three - in which Groefsema's analysis of *so* will be discussed - I shall briefly outline two of the consequences of the above reformulation of the conceptual/procedural distinction in line with Groefsema's arguments. Firstly, the reformulation entails that, in principle, there is no reason why a single expression could not encode both conceptual and procedural information, since all stem from the same set of ontological categories. This possibility will be explored in relation to the identification of procedural encoding constraining phase two of utterance interpretation. Secondly, Groefsema suggests that not only propositions, but other less complex (i.e. sub-propositional) conceptual representations may be amenable to manipulation on the basis of linguistically encoded constraints on interpretation; she writes:

"we do not have to postulate that there is a separate 'procedural theory' to account for the interpretation of discourse connectives. A consequence of this is that we end up with a unified account of constraints on interpretation: constraints may range over the kind of logical form that can be built, they may range over the interpretation of elements within a logical form..., or they may range over whole propositions, as is the case with *so*."

(Groefsema 1992:220)

This allows for the possibility of procedural information operating not only at phase three, constraining the inferences to be drawn from propositional conceptual representations, and at phase two, constraining the process of inferential enrichment of semantic representations (see §2.1.1), but also at phase one, constraining "the kind of logical form that can be built". If the informational encapsulation of the language

module is to be preserved (i.e. if the modularity hypothesis is to be retained) procedural information can operate at phase one only in so far as it imposes constraints on anticipatory logical hypotheses, thereby disposing amongst alternative logical forms and avoiding possible subsequent ambiguity; that is, procedural encoding at phase one, if it exists, will have an inhibitory effect only on the decoding process. However, phase one of utterance interpretation, being a fast and efficient modular process, involves relatively little processing effort in comparison with the inferential phases, and it is debatable whether procedural information at this phase, however it might function, would reduce overall processing effort. The possibility of procedural encoding at phase one will not be addressed here; instead I shall concentrate on procedural encoding at phases two and three.

## **2.2 Exemplification of Procedural Encoding at Phase Three**

### **2.2.0 Introduction**

In this section I shall attempt to clarify the nature of procedural encoding at phase three with reference to four examples. I shall first review and extend accounts of two discourse connectives as exponents of procedural encoding: Blakemore's (1987, 1988a) account of *so* and Jucker's (1993) account of *well*. Second, whilst there have been many relevance theoretic studies of discourse connectives (not just in English) which treat them as encoding procedural constraints on the computation of implicatures (cf. Luscher (1994) and references therein), there have been far fewer studies addressing non-lexically encoded information; the discussion of Vallduví's (1993) theory of information packaging in §2.2.2 attempts to demonstrate how certain aspects of word order and intonation can also be accounted for in terms of procedural encoding constraining the computation of implicatures.

Conversely, it has been suggested, with respect to the computation of higher-level explicatures, that non-declarative word order in English is an example of procedural encoding constraining the way in which a speaker intends a proposition to achieve relevance (Wilson & Sperber 1993a:22). The suggestion is that the proposition is 'mentioned' not as an assertion but in order to seek additional information about it, such as confirmation or disconfirmation of its truth. §2.2.3 suggests that procedural encoding of this sort can be lexical as well as syntactic and intonational, and takes the Kiswahili interrogative markers *je* and *-je* as an example.

### 2.2.1 Constraints on implicatures 1: Discourse connectives *so* and *well*

*So*

In this subsection I shall review Blakemore's (1987, 1988a) procedural account of *so*. This begins with a discussion of Grice's brief comments on the role of *therefore* in (2.5):

(2.5) He is an Englishman; he is, therefore, brave.

Grice claimed that *therefore* in (2.5) indicates that his being brave is a consequence of his being an Englishman. However, the term *consequence* can indicate either causal effect:

(2.6) He is an Englishman. Because of that he is brave.

[Blakemore 1987 (20)]

or a logical conclusion - i.e. the fact that he is an Englishman is evidence for the belief that he is brave. "In other words," asks Blakemore (1987:78), "does *therefore* indicate a causal relation between states of affairs or an inferential relation between propositions?" Because *therefore* in (2.5) is parenthetical, that is, it is intonationally separated from the rest of the utterance, as indicated by the commas, Blakemore (1987:80) claims that *therefore* in (2.5) expresses an inferential relation; (2.5) does not mean (2.6). It is perhaps best not to read too much into Grice's use of parenthesis, however, it is still true to claim that (2.5) does not mean (2.6), if "mean" indicates semantic identity. Thus (2.5) could (minus parenthesis) be used to express the proposition in (2.6), or it could express that his being an Englishman is reason to believe that he is brave.

In (2.7) and (2.8) on the other hand, only an inferential interpretation is possible:

(2.7) She's your teacher. Therefore respect her.

(2.8) She's your teacher. Therefore you must respect her.

[Blakemore 1987 (23) and (24)]

Blakemore (1988a:186) notes that although *therefore* does not contribute to the truth-conditions of utterances that contain it, it does affect the way propositions are interpreted. In (2.5) for example, *therefore* indicates that the proposition [He is an Englishman] is relevant because it contributes evidence for the proposition [He is



brave]<sup>22</sup>. In other words, *therefore* provides addressees with an indication of how utterances containing it are to be inferentially interpreted in line with the principle of relevance. The least effort clause of the principle of relevance justifies the existence of such expressions, since by constraining the utterance interpretation process, *therefore* reduces the processing effort required to compute the relevance of an utterance containing it.

*So* behaves similarly to *therefore*, but with some striking differences. Blakemore claims that *so* could replace *therefore* in all the above examples without infelicity except that *therefore* but not *so* can be used to express a causal connection. If *so* can be used to express causal connections it should always be possible to substitute it for expressions such as *because of that* and *as a result* without infelicity. Blakemore (1987:88, 1988a:192) claims that (2.10) below is acceptable only to a hearer who already assumed that anyone who ate the condemned meat would fall ill thirteen hours later (in which case *so* could indicate an inferential rather than a causal relation), whereas the acceptability of (2.9) does not depend on this prior assumption:

- (2.9) Tom ate the condemned meat. Because of that / As a result he fell ill thirteen hours later.
- (2.10) Tom ate the condemned meat. So he fell ill thirteen hours later.
- [Blakemore 1987 (40) and (41)]

However, native speakers of British English (none of whom already entertained the assumption that eating condemned meat results in illness thirteen hours later) to whom (2.10) was presented have found it perfectly acceptable (and also where *so* has been replaced with *therefore*); they generally commented that *thirteen hours later* was incidental - the fact that Tom fell ill being the salient point. *So* can be used in other expressions in which the proposition it introduces is not already held as an assumption by the addressee and in which the relation between the states of affairs is a causal one. For example (2.11) below, describes a causal relation:

- (2.11) The child survived largely on a diet of maize and so developed kwashiorkor.

I assume that most readers will be unfamiliar with kwashiorkor and yet still find (2.11) acceptable (if obscure). This demonstrates that *so* is not only acceptable when the addressee is assumed to already entertain the assumption that two conjoined propositions are causally linked in a particular way (and will hence form an

---

<sup>22</sup> These are strictly only propositions once the referent of 'He' has been determined. Also, Blakemore (1992:153 fn.16) concedes that *therefore* can be truth-conditional and conceptual if (2.5) is interpreted as 'It is a consequence of his being an Englishman that he is brave.'

inferential connection). The relevance of (2.11) above lies in the assumption that the child's kwashiorkor was caused by diet. The relevance of the more explicit (2.12), which defines kwashiorkor, lies in the implication that maize is not high in protein, which the addressee may or may not be assumed to entertain as an assumption, (a further highly relevant implication might be that a high protein diet will improve the child's health).

(2.12) The child survived largely on a diet of maize and so developed kwashiorkor which is a wasting disease resulting from protein deficiency.

Awareness that the link between the states of affairs, 'The child survived largely on a diet of maize' and 'The child developed kwashiorkor' is one of causal effect, arises from determining the contextual effects of an utterance of (2.11) on the basis of (among other factors) world knowledge. The relation between *representations* of these states of affairs, on the other hand, as encoded by *so*, is less precise, and consists simply of the instruction to make an (unspecified) inferential connection.

Whilst the ability to express causal relations does not distinguish *so* and *therefore*, genuine differences between these forms can be seen in the following examples:

(2.13) There's \$5 in my wallet. So I didn't spend all my money then.

(2.14) There's \$5 in my wallet. ??Therefore I didn't spend all my money then.

[Blakemore 1988a (11) and (12)]

"The addition of sentence final *then* in these examples indicates that the second proposition must be construed as the specification of the significance or relevance of the first proposition rather than as a proposition whose truth is proven by the first proposition. In other words, *so*, but not *therefore* can be used to indicate that the relevance of the proposition it introduces lies in the fact that it is a contextual implication of the first proposition."

(Blakemore 1988a:188)

The claim that, "*so*, but not *therefore* can be used to indicate that the relevance of the proposition it introduces lies in the fact that it is a contextual implication of the first proposition," does not conflict with my arguments above that *so* can express causal connections. Contextual implications can arise when either a causal or an inferential relation is assumed to hold between two propositions. Expression of either type of relation is not part of the semantics of *so*; however an utterance containing *so* can be inferentially enriched to give rise to either of these readings.

In addition to indicating that one proposition is a contextual implication of another, *so* can be used to confirm the relevance of a remark by checking that the intended implication has in fact been drawn:

- (2.15) A: Tom's car isn't here.  
B: So he decided not to come after all. [Blakemore 1987 (34)]

Or if the intended contextual implication (and hence the relevance) of a proposition is unclear, the response, 'So?' or 'So what?' may be appropriate, whereas 'Therefore?' is not.

One final difference between *so* and *therefore* cited by Blakemore is that *so* but not *therefore* can be used when there is no antecedent proposition. If the addressee arrives home laden with parcels, (2.16) but not (2.17) would be an appropriate response:

- (2.16) So, you've spent all your money.  
(2.17) ? Therefore, you've spent all your money.  
[Blakemore 1988a (13) and (14)]

The preceding examples and analyses suggest that the semantics of *so* consists of an instruction to the hearer to establish an inferential connection with some mutually manifest assumption (not necessarily derived from a preceding utterance as we have seen). Although a conceptual representation derived from an utterance of a sentence containing *so* may contain an assumption functionally equivalent to the logical operator  $\supset$ , this is inferred rather than encoded by *so*.

*So* has other discourse functions which can be accounted for by the above characterisation of its semantics. Firstly, *so* often marks a change of topic or a return to a previous topic. For example,

- (2.18) So, what shall we have for dinner?

uttered 'out of the blue' is felicitous when something in the utterance interpretation context makes this a relevant remark; in fact (2.18) was uttered at about six o'clock when dinner is usually prepared (all the following examples were noted in authentic conversation). The utterance of (2.18) served a dual purpose: it posed a worthwhile question and it drew the addressee's attention to the fact that the underlying proposition, 'we shall have dinner', was inferable from the context; that is, it

prompted the addressee to pay attention to the context, specifically the time. In Blakemore's example,

(2.16)           So, you've spent all your money.

the communicative intention could be as much to draw the addressee's attention to the number of parcels they are carrying and to the fact that the speaker has noticed this, as it is to convey the proposition expressed. The speaker of (2.16) may well not be concerned with its truth; an utterance of (2.16), as of (2.18), is a device for making some mutually manifest assumption more manifest.

*So* can also be used to mark a return to a previous topic, like 'as I was saying' with which it can co-occur. *So* makes explicit the fact that there is some inferential connection between the context and the content of the following utterance, and can direct the addressee to view the content of the forthcoming utterance as inferable from some mutually manifest assumption. Related to this is the use of *so* in phatic communication, where it can precede a representation of an event which is not merely inferable from some mutually manifest assumption but is identical with one:

(2.19)           So, here we are.

This conveys no new information, but merely makes explicit some mutually manifest assumption. *So* is also commonly used in introductions, for example at parties, and in interviews and supervisory sessions:

(2.20)           So, how do you know Nickie?

(2.21)           So, how's the work going?

Here, *so* introduces a new topic which is in some way derivative from the context. In (2.20) both the interlocutors were unknown to each other but met at Nickie's party, therefore it can be assumed that both know Nickie. In (2.21) the mutually manifest assumption is that the interlocutors have met for a supervisory session in which the addressee's work will be discussed.

The above paraphrase of the semantics of *so* can be augmented to take account of this retrospective effect: *The semantics of so consists of an instruction to the hearer to establish an inferential connection with some mutually manifest assumption (not necessarily derived from a preceding utterance) to which it may in turn draw the addressee's attention.*

I shall now compare the above paraphrase of the semantics of *so* with Groefsema's (1992) analysis of *so* as encoding a concept which gives access to a particular logical selection frame. A successful reformulation of content of this selection frame (as an example) will lend support my adoption of a number of Groefsema's arguments in §2.1.2.

Like the definite and indefinite articles, the discourse connective *so* does not belong to any conceptual category according to Groefsema (1992), and neither can it combine with anything to yield a well formed formula in the language of thought. There is therefore only one tier (the top tier) in Groefsema's (1992:220) representation of *so*, since the lower tier contains no information:

[PROPOSITION]  
logical consequence of a proposition in context

There are three faults with Groefsema's representation of *so*: First, Groefsema (1992:220) claims that "by necessity *so* introduces a proposition" which is indicated in the logical selection frame of the concept *SO* as [PROPOSITION]. This is clearly not the case as the examples (2.7) and (2.18) illustrate:

(2.7) b. She's your teacher, so respect her.

(2.18) So, what shall we have for dinner?

In (2.7b) *so* introduces an imperative and in (2.18) it introduces a *wh*-interrogative. In these examples *so* could be viewed as introducing a complex proposition (as opposed to more than one conjoined propositions) if the Davidsonian approach to mood and performance is adopted, although there is no evidence that this is (consciously) the case. In the speech act literature (e.g. Davidson 1979), utterances of non-declarative sentences have been analysed as consisting of the performance of two speech acts; a non-declarative sentence consists of a declarative core plus a 'mood-setter'. Both the declarative core and the mood-setter have a truth value (unlike the combined utterance, since this is not the utterance of a conjunction), and may therefore have the status of propositions. However, this entails that, in relevance theoretic terms, the propositional attitude of an utterance is as much a proposition (albeit a dependant one) as the propositional form itself; in this case the distinction, which Groefsema (1992) consistently makes, between logical form (which can be enriched into a propositional form) and semantic representations other than logical form (including those elements which can be enriched into a propositional attitude) must be vacuous.

Second, the logical selection frame of *SO* contains the information that the proposition introduced by *so* is a logical consequence of a proposition in context; as we have

seen, however, one of the differences between *so* and *therefore* is that *so* but not *therefore* can be used when there is no overt antecedent proposition (for example in response to the addressee arriving home laden with parcels):

(2.16)            So, you've spent all your money.

(2.17)            ? Therefore, you've spent all your money.

This leads to the third inadequacy of Groefsema's representation of *so* which is that, as it stands, the logical selection frame proposed for SO could, if its inaccuracies were overlooked, serve equally well as the logical selection frame of THEREFORE. In other words, it fails to capture the differences between *so* and *therefore*, and suggests that both *so* and *therefore* map onto the same concept. To avoid these inadequacies, the logical selection frame for SO could be reformulated as follows:

[[PROPOSITION] ([PROPOSITIONAL ATTITUDE])]  
inferentially connected with some mutually manifest assumption

The fact that an instance of *so* gives rise to the above logical selection frame also accounts for the fact that *so* may draw the addressee's attention to some mutually manifest assumption, since explicit mention is made of some such assumption in the logical selection frame itself.

### *Well*<sup>23</sup>

According to Jucker's (1993) procedural analysis, the presence of the discourse connective *well* in an utterance serves to help the addressee select the intended interpretation context. It does this in a negative way, by indicating that the currently most accessible context is inappropriate, thus saving the addressee the effort of searching in vain for contextual effects in that context. The aspects of the utterance interpretation context to which *well* draws the addressee's attention are assumptions made on the basis of previous discourse. Jucker describes *well* as,

"a signpost signalling to the hearer that the context created by the previous utterance - whether produced by the current speaker or the current hearer - is not the most relevant one for the interpretation of the impending utterance."

(Jucker 1993:440)

Unsurprisingly, therefore, *well* is commonly used by politicians in radio and television interviews. Jucker (1993) identifies four distinct uses of *well* which will be exemplified in turn (all examples are from The World at One, BBC Radio 4, 22/9/95):

---

<sup>23</sup> An earlier version of this sub-section appeared as part of Nicolle (1995).

i) as a marker of insufficiency, for instance in response to an interrogative where *yes/no* is not a sufficient response, or where neither straightforward agreement nor denial is an appropriate response to a statement, as in the reply given by David Davis (British Foreign Office Minister) to A's question in (2.22):

(2.22) A: ...now that surely is not being at the heart of Europe that is, as Robin Cook says, an acceptance sometimes of marginalisation ... if it is in Britain's interest.

DD: Well, you may view it as marginalisation ... I. let me give you an example...

(2.23) A: What um was Mr. Major's mood when he arrived there is he is he looking forward to this or is he think he's going to get as ... tough a time as he usually gets at these mini summits?

B: Well he seemed in a very bouncy mood...

The use of *well* in (2.23) can be analysed as a marker of insufficiency for various reasons. First, A asks three separate questions of B, so clearly a straightforward *yes* or *no* response is insufficient. Also, if B is taken as responding to the final question about whether Mr. Major thinks he is going to have a tough time, the use of *well* indicates that B doesn't know what Mr. Major thinks (he has insufficient knowledge to answer this question) but from appearances the affirmative response which A seems to expect is mistaken.

ii) as a face-threat mitigator, by which Jucker means that the presence of *well* indicates a problem on the interpersonal level in which either interlocutor risks losing face should their background assumptions be contradicted. In (2.24) Frank Dobson (Labour's local government spokesman) risks losing face if he accepts the suggestion that certain aspects of his party's latest policy document have not been thought through:

(2.24) A: what about these hit squads who would go in if there is - it sounds rather like a broken marriage - if there is an 'irretrievable breakdown in' I mean who decides if it is [broken down]

FD: [w... w... w... ] well it's set out i. ... it's set out very clearly in the document.

iii) as a frame, separating discourse units, in particular at the beginning of a turn, or to introduce direct reported speech (note that in (2.25) *well* indicates a new topic, but occurs, not at the beginning of the turn, but after a short preamble which sets the scene for the topic without adequately introducing it):

(2.25) A: ...the Tories lose no opportunity of pointing to either financial fecklessness or political correctness in local Labour-run councils. Well today Labour has unveiled its plans for reinvigorating local government...

(2.26) B: I asked him specifically about that speech and how it should be interpreted by ... Europe's leaders and he said well that's exactly the same sort of thing that I was trying to say at er Leiden.

iv) as a delay device. Jucker's example and comment are given:

(2.27) R: ...well never mind. It's not important.  
D: Well, it is important.

(Pomerantz 1987: 74)

In this use there is some overlap with the category of *well* as a face-threat mitigator because here, too, *well* indicates problems on the interpersonal level."

(Jucker 1993:448)

Jucker's analysis is essentially accurate, but can be refined in two respects. First, intonation is crucial to the interpretation of *well*. Jucker only mentions prosody once, in his discussion of *well* as a frame, and then only to dismiss it as largely irrelevant:

"If the words in reported direct speech cannot be claimed to be verbatim, it makes little sense to distinguish between those instances of *well* that belong to the reporting clause and those that belong to the quotation. The original words may or may not have included *well*, even if the prosody of the reporting clause and the quotation seem to indicate a difference between the two."

(Jucker 1993:447)

Prosodically, *well*, like *so* and other discourse connectives, is parenthetical; a number of distinct prosodic forms can be distinguished in utterances of *well* in discourse. As a marker of insufficiency, *well* is typically lengthened and at a steady, moderately high pitch level; as a face-threat mitigator, *well* is again typically lengthened but with a rising-falling pitch; and as a frame, it is significantly shorter in duration than in its other uses and may occur at various pitch levels. As a delay device, there is considerable prosodic variation and any of the above three prosodic forms are possible, which leads to my second refinement of Jucker's analysis: that the use of *well* as a delay device can invariably be subsumed under one of the other three uses.

An utterance of *well* as a discourse marker, in whatever use, indicates, (a) an intention to speak, and (b) that the most accessible background assumptions are insufficient for the interpretation of the impending utterance. Hearers will therefore assume that there is no point in continuing or initiating their own turn until the background assumptions have been renegotiated. A delay may then follow in order to give the interlocutors time to reorient their assumptions - the new addressee with regard to what aspect of



the current context is irrelevant for the interpretation of the impending utterance (based on an analysis of prosody), and the new speaker with regard to what background assumptions can be mutually assumed, how to tactfully save face, etc.

This analysis could be falsified if it could be shown that *well* is ever used in situations where the most accessible background assumptions are clearly appropriate and sufficient and where this was recognised by the speaker. Consider the following (invented) dialogue:

- (2.28) A:     Could I borrow your pen, please?  
      B:     Well, yes.

B's response would be felicitous if *well* were used as a marker of insufficiency ('Well, yes, but it's not working very well'), as a face-threat mitigator ('Well, yes, of course you can - you don't have to ask'), or as a frame ('Well, yes, I've been meaning to have a discussion with you about pens...'). However, where none of these situations pertain, the use of *well* is infelicitous, and so the procedural analysis of *well* is not falsified.

In summary, then, what is common to all the uses of *well* is the function of indicating that the currently most salient context is deficient in some way for the correct interpretation of the forthcoming utterance. The semantics of the discourse connective *well* is unitary, but underdetermined; intonation is one indication of the way in which *well* is to be inferentially enriched to specify more precisely the nature of the contextual deficiency. (Thereby demonstrating that semantic representations of linguistically encoded elements outside of logical form can also undergo inferential enrichment.)

### *Conclusion*

I began my investigation into *so* and *well* with the claim that they both encode procedural information constraining phase three of utterance interpretation, the output of which is a contextual effect (or effects). Specifically, they perform this function by stipulating constraints on the identification of implicatures; implicatures are assumptions which are communicated by the use of a particular utterance without being developments of the logical form encoded by that utterance. In §1.4.1, contextual effects were defined as assumptions arising out of the union of 'old information' {C}, that is, currently held assumptions, and 'new information' {P}, typically assumptions ostensibly communicated by a speaker. It is the 'contextualisation' of {P} in {C}, that is, the way in which {P} is related to {C}, that

determines which contextual effect(s) are computed by the addressee. I have demonstrated that what discourse connectives do is constrain the contextualisation of {P} in {C}. The use of the term 'discourse connective' to designate *so* and *well* is somewhat misleading, since it suggests that what *so* and *well* connect is stretches of discourse. As we have seen, this is not always the case (see for example (2.16)); there are many ways in which prior assumptions {C} can come to be held, and previous discourse is only one of them.

The contextualisation of {P} in {C} is constrained in the following different (and contradictory) ways by *so* and *well*. *So* instructs the addressee to establish an inferential connection between {P} and {C}, and in the process may guide the addressee to the particular mutually manifest assumption(s) of which {C} consists. *Well*, on the other hand, indicates that the speaker's intended contextual effects will not be computed if {P} is contextualised in {C}, where {C} consists of the currently most accessible background assumptions. That is, it signals that {C} is inappropriate, thus indicating that the addressee should select an alternative set of background assumptions {C'} in which to contextualise {P}. Intonational differences provide some indication as to the nature of the inappropriateness of {C}.

Some of the communicative functions of both *so* and *well* are illustrated in the following (authentic) exchange, which will be analysed in accordance with the above characterisations of *so* and *well*:

- (2.29) A (a teacher):           We're having a non-uniform day at school tomorrow.  
      B (A's husband):        What, for the staff as well?  
      A:                         Yes.  
                                  [*gap of approximately ten minutes*]  
      B:                         So, what are you going to wear tomorrow?  
      A:                         Well... I don't know.

B's use of *so* helps A to determine the intended contextual effects of his utterance in the following way: *So* indicates that there is an inferential connection between {P} (the assumptions derivable from the question "what are you going to wear tomorrow?") and some mutually manifest assumption or set of assumptions {C}; this prompts A to look for some mutually manifest assumption on the basis of which {P} could be inferred. This assumption was made mutually manifest some ten minutes previously through A's utterance "We're having a non-uniform day at school tomorrow" from which it can be inferred that A has to make a decision about what to wear. A's use of *well* in response to B's question indicates that the currently most

accessible assumption {C'} is inappropriate and hence A's response {P'} will fail to achieve the intended contextual effects if contextualised in {C'}. The assumption that A has decided what to wear (inferred from the use of *...going to wear* - see §6.4) is inappropriate because false, hence A uses *well* as a marker of insufficiency.

### 2.2.2 Constraints on implicatures 2: Information packaging

Thus far, I have demonstrated that procedural information constraining the identification of implicatures can be lexically encoded in the form of discourse connectives. I shall now investigate whether, and in what ways, such procedural information might be encoded syntactically. It has been suggested (Wilson & Sperber 1993a) that differences in word order can (in English) encode procedural information which contributes to higher-level explicatures through differences in mood (indicated by the respective ordering of the grammatical subject and finite verb of a clause). Differences in word order can also constrain the proposition expressed:

Mary loves Peter      vs.      Peter loves Mary

At other times, differences in word order merely guide the search for intended contexts and contextual effects without altering either the proposition or the higher level explicatures associated with unmarked declarative word order. Examples include:

topicalisation:	She loves Peter	vs.	Peter she loves
passives:	Peter is loved by Mary		
clefts:	It's Peter that she loves		
pseudo-clefts:	Who she loves is Peter, etc.		

The effect of such differences in word order is to specify the nature of the relation between the associated proposition and the utterance interpretation context; in other words, to guide the addressee as to how {P} is to be contextualised in {C}. Similar effects can be achieved in many languages through variations in intonation. Thus, word order can contribute both conceptual and procedural information. This fact is reflected in functional grammar, with its long established three-way division into experiential meaning (propositional), interpersonal meaning (syntactic mood and illocutionary force: higher-level explicatures in relevance theoretic terms) and textual meaning (which excludes higher-level explicatures and is characterised syntactically as theme-rheme).

What functional grammarians call textual meaning is developed in Vallduví's (1993, 1994) theory of information packaging. Whilst Vallduví's theory of information packaging is neither a branch of functional grammar nor a form of relevance theory (it is, rather, a version of update semantics incorporating ideas drawn from Heim's File Change Semantics, and, like relevance theory, assumes a generative perspective) it is discussed here since most relevance theoretic accounts of procedural constraints on implicatures have followed Blakemore's lead in dealing with lexical rather than syntactic procedural encoding. I shall demonstrate that, whilst Vallduví (1993) provides a detailed analysis of information packaging in English and Catalan in particular, the principle of relevance provides a complementary account of why interlocutors exploit information packaging resources.

Information packaging is a pragmatic notion concerning the processing of information, where 'information' is defined by Vallduví as that part of the propositional content of a sentence which contributes to the hearer's knowledge-store. Information packaging consists of a small set of linguistically encoded instructions with which a speaker instructs a hearer to retrieve the information carried by a sentence and enter it into her/his knowledge-store in a specific way (Vallduví 1993:14-18). In relevance theoretic terms, information packaging involves constraints on how new assumptions are intended to be added to existing ones - i.e. it constrains the addressee's search for intended contextual effects. Specifically, information packaging guides the addressee towards those subsets of the communicated assumptions which constitute {P} and {C}, such that the contextualisation of {P} in {C} will result in adequate contextual effects at the cost of minimal processing effort. I shall address this relevance theoretic interpretation after a brief discussion of information packaging itself.

Vallduví's theory of information packaging is restricted in that information packaging does not belong to the same class as illocutionary acts (Vallduví 1993:30-31) or Gricean implicature (ibid. 31-33), both of which are subsumed under relevance theory. Furthermore, Sperber & Wilson's discussion of "stylistic effects" (S&W 1986:202), roughly equivalent in scope to the textual part of meaning in functional grammar, subsumes both information packaging considerations and cohesion (Halliday & Hasan 1976). There is a high degree of compatibility between the theories however. Both theories reject such distinctions as theme-rheme, topic-focus, given-new, as adequate to a linguistic description of sentences (Vallduví 1993:35-53; S&W 1986:210-217), but recognise that utterances of sentences can contain more informative and less informative parts (Vallduví 1993:35) corresponding to foreground and background, which "arise as automatic effects of the hearer's

tendency to maximise relevance, and of the speaker's exploitation of that tendency" (S&W 1986:217).

Whilst Vallduví describes in some detail the S-structure realisation of information packaging, Sperber & Wilson merely refer to such a possibility in a footnote (1986:263 n.21), and subsequent relevance theoretic studies have concentrated on lexically encoded procedural indicators such as *well* and *so* rather than on procedural syntax. Vallduví posits a distinction between LF (logical form in GB) and information structure (IS) whereby IS is derived from S-structure in the same way that LF is, according to transformational rules. In the same way that LF links S-structure and logical semantics, IS links S-structure and informatics (Vallduví 1993:142). Vallduví (1993:54-65) characterises the S-structure realisation of IS in the following terms:

$$S = \{\text{FOCUS, GROUND}\} \quad \text{GROUND} = \{\text{TAIL, LINK}\}$$

The FOCUS is "the only nonelidable part of the sentence, since it is the only contribution to the hearer's knowledge-store at the time of utterance (or so the speaker assumes)" (Vallduví 1993:57-8). In relevance theoretic terms, the FOCUS can be viewed as the new information contained in {P}. The LINK is the "address pointer" and - obligatorily in Catalan or, in English, when not prosodically marked - must be sentence-initial, following Halliday's (1967) restriction on themes. In the same way that there can be multiple themes, there can be more than one LINK in a sentence (a "link string"). The TAIL is what remains. I shall view the LINK and the TAIL as specifying that subset of {C} in which a contextualisation of {P} is intended to achieve adequate contextual effects. The LINK and TAIL both express 'background information' in the sense of S&W (1986:202-217) and are informationally more alike than distinct; indeed Vesselin Vatchkov (p.c. 9/6/96) has suggested - on account of cross-linguistic data from 'pragmatically ordered languages' (Mithun 1992) such as Cayuga (Iroquoian), Ngandi (Australian) and Coos (Penutian, Oregon) which present focal information first (and therefore have no LINKs), and languages like Czech which typically present focal information last (and therefore rarely exhibit TAILS) - that the LINK/TAIL distinction be abandoned in favour of some kind of 'split topic' (or split GROUND) analysis. Whilst bearing these considerations in mind, I shall maintain Vallduví's terminology as it provides a convenient representation of IS in English, Catalan and Kiswahili - the languages discussed here.

On the basis of English and Catalan data, Vallduví (1993:62-65) identifies four possible "formational structures" that a sentence can have; they are: LINK-FOCUS;



creamy sauce was on the menu some years ago, this would make assumptions 1 and 2 manifest to the addressee(s) and would allow the addressee(s) to infer assumptions 3, 4, 5, etc. depending on their individual background knowledge:

1. The diners will be eating broccoli;
2. the broccoli will be served in a creamy sauce;
3. the diners like broccoli in a creamy sauce;
4. the President, George Bush, will be among the diners;
5. the First Lady will be among the diners, etc.

These are all possible assumptions against which an utterance (by the head waiter) of (2.30a) could be processed:

(2.30) a.      The boss      HATES      broccoli.  
                 LINK          FOCUS      TAIL

The response in (2.30a) achieves optimal relevance in the following way: The FOCUS 'HATES' specifies new information which is neither manifest in the chef's announcement, nor inferable from it, and is therefore prosodically marked (Vallduvf's use of capitals indicates this). Identification of the FOCUS activates a focal scale of entailments which includes the background entailment 'the President feels some way about broccoli' (combining the LINK and the TAIL). This background assumption results from the following chain of inferences: (i) combine assumptions 3 and 4 by modus ponens, giving 'the President likes broccoli in a creamy sauce', (ii) trivially infer 'the President likes broccoli', (iii) trivially infer 'the President feels some way about broccoli'. This chain of background entailments constitutes {C}, the subset of existing assumptions against which (2.30a) is intended to be processed. However, the assumption 'the President likes broccoli' is contradicted by the proposition expressed by (2.30a), so (2.30a) achieves relevance by eliminating this assumption from {C}.

Had 'The boss' been the FOCUS, it would have given rise to a chain of entailments including 'somebody hates broccoli'; since this is not a background assumption, making 'the boss' the FOCUS in this exchange would have been infelicitous. Similarly, making 'broccoli' the FOCUS would have given rise to a chain of entailments including 'the boss hates something' which again is not a background assumption, and would have resulted in infelicity.

We now turn to the ALL-FOCUS constructions:

- (2.31) {C} *What doesn't the boss like?*  
 {P} BROCCOLI.
- (2.32) {C} *How does the boss feel about broccoli?*  
 {P} He HATES it.

In (2.31) {C} consists of the assumption that the boss doesn't like something plus the higher level explicature that the speaker would like to know what this is; {P}, the response 'BROCCOLI', thus achieves adequate contextual effects (it specifies what the boss doesn't like) for minimal processing effort. In (2.32) {C} consists of the assumption that the boss entertains some feeling about broccoli plus the higher level explicature that the speaker would like to know what this is; {P}, the response 'He HATES it' achieves adequate contextual effects (it specifies how the boss feels about broccoli) whilst putting the addressee to minimal processing effort. The pronouns 'he' and 'it' are grammatical requirements since English generally requires an overt grammatical subject in each clause and the verb *hate* subcategorises for an object.

Information packaging considerations and the relevance theoretic justification for them outlined above hold also for Kiswahili. Variations in word order are used more frequently in Kiswahili than in English to indicate procedural information and associated contextual effects such as implicit contrast and emphasis. Reasons for this include the lack of a syntactic mood function in Kiswahili to constrain higher-level explicatures (as noted in §2.2.3 below), and the fact that Kiswahili speakers often employ syntactic and morphological devices where English speakers might use intonation. The following examples demonstrate how word order contributes to procedural meaning at phase three in Kiswahili. Examples (2.33) to (2.38), below, all refer to the same event; all convey precisely the same conceptual information, except (2.38) which omits mention of agency. Differences in procedural information will be described by reference to Vallduví's classification.

[Key: SP = 'subject prefix'; OM = 'object marker'; subscripted numbers refer to noun class (final -a is the indicative suffix). More detailed information on Kiswahili is provided in §3.1.3.]

- (2.33) A: *What did you do?*  
 B: ni + li + vunja kikombe  
 SP PAST break cup<sub>7</sub>  
*I broke a cup*
- formational structure ALL-FOCUS



Since (2.33) is discourse initial it can make no reference to any assumption established by previous discourse; all of the information is new information, i.e. {P}, and so (2.33) is ALL-FOCUS.

- (2.34)           A:    *What did you do to the cup?*  
                   B:    ni + li + ki + vunja  
                           SP PAST OM7 break  
                           *I broke it*  
 formational structure    ALL-FOCUS

In (2.34), the morphologically bound pronouns *ni-* and *-ki-* (SP and OP respectively) are functioning like the pronouns in (2.32a); both require a minimal amount of processing effort and refer to highly accessible antecedents and therefore neither is a LINK or a TAIL.

- (2.35)           A:    *What happened to the cup?*  
                   B:    (kikombe) ni + li + ki + vunja  
                           cup7            SP PAST OM7 break  
                           (*the cup*)    *I broke it*  
 formational structure    (LINK)            FOCUS

- (2.36)           A:    *What happened to the cup?*  
                   B:    ni + li + ki + vunja (kikombe)  
                           SP PAST OM7 break cup7  
                           *I broke it (the cup)*  
 formational structure    FOCUS                (TAIL)

In both (2.35) and (2.36) the VP *nilikivunja* (I broke it) is the FOCUS since it constitutes the new information, {P}, to be contextualised in {C} which contains the assumption 'something happened to the cup'. The use of *kikombe* (the cup) as a LINK or TAIL is informationally superfluous (it is obvious that the cup is at issue) but by directing A towards this address nonetheless, B topicalises *kikombe* (for example the cup might have been faulty, or the cup was broken in contrast to, say, the teapot when both were dropped).

- (2.37)           A:    *Who broke the cup?*  
                   B:    (kikombe) ki + li + vunj + w + a    na mimi  
                           cup7            SP7 PAST break PASSIVE by me  
                           (*the cup*)    *was broken by me*  
 formational structure    (LINK)            LINK                FOCUS

- (2.38)           A:    *What happened to the cup?*  
                   B:    (kikombe) ki + me + vunj + ik + a  
                           cup<sub>7</sub>            SP7 PERF break STATIVE  
                           (*the cup*)       *broke*  
 formational structure       (LINK)            FOCUS  
                           (with the stative suffix there can be no reference to an agent)

Finally, (2.37) and (2.38) differ primarily over whether agency is mentioned. In both, *kikombe* has been made the grammatical subject but whilst (2.37) provides the information that it was B who broke the cup, and hence *mimi* (me) is the FOCUS, (2.38) omits mention of the agent (using the stative rather than the passive suffix) and makes the VP *kimevunjika* ('it broke') the FOCUS, thereby excluding the question of agency (and hence guilt) from the contextual effects which A is licensed to deduce.

Although, the propositional form of (2.33) to (2.37) is the same, the procedural information encoded by the variations in syntax and agreement, together with passive and stative morphology in (2.37) and (2.38) respectively, imposes various constraints on the processing of utterances of these examples. For example, B's response in (2.37) would be inappropriate in response to the question *What did you do?* The LINK or TAIL, when realised, must refer back to a relevant constituent of the utterance interpretation context {C}, and the FOCUS should identify the new information {P} which the addressee is expected to add to {C}.

### *Summary*

Information packaging reduces the addressee's processing effort by indicating, through systematic variations in word order and intonation, those subsets of the communicated assumptions which constitute {P} and {C}. Speakers do not use the information packaging resources of their language to provide additional information, rather they use information packaging to present information in such a way as to facilitate the contextualisation of new information {P} in existing assumptions {C}. The function of a LINK or a TAIL is to specify the subset of existing assumptions which constitutes {C}; where the relevant subset of existing assumptions is mutually manifest, the speaker does not need to provide a LINK or a TAIL and may use, say, a pronoun rather than a definite description. The FOCUS specifies the new information {P} and is non-elidable and may be intonationally prominent. Taken as a whole, information packaging has the effect of reducing the effort required by the addressee to contextualise {P} in {C}, and can therefore be classified as an example of procedural encoding at phase three constraining the computation of implicatures.

### 2.2.3 Constraints on higher-level explicatures: Kiswahili *je/-je* interrogatives

So far I have discussed procedural encoding at phase three constraining the computation of implicatures. Such encoding can be both lexical, as in the case of the discourse connectives *so* and *well*, and syntactic, in the case of information packaging. We have also seen that intonation has a role to play, and could also be viewed as encoding procedural information. We now turn to the second type of procedural encoding at phase three (first mentioned in §2.1.1) - procedural encoding which constrains the interpretation of higher-level explicatures (introduced in §1.3.3) which indicate propositional attitudes and speech act information. In §2.1.1 I noted Wilson & Sperber's (1993a) claim that differences in mood in English encode procedural information, the purpose of which is to constrain the identification of higher-level explicatures. Such information comes in various forms. Wh-questions present the addressee with an incomplete proposition and an indication as to how the speaker would like it completed, encoded by a particular interrogative expression. Polar interrogatives (yes-no questions) may be indicated by changes in word order, distinctive intonation, interrogative particles, or a combination of these. This information may also be conveyed intonationally in English. As an example of lexical rather than syntactic procedural encoding constraining higher-level explicatures, I have selected the Kiswahili interrogative markers *je* and *-je*.

Ashton (1947:151) lists three ways in which questions can be formed in Kiswahili; these are: 1) by means of a statement with declarative word order but distinctive intonation, sometimes introduced (or followed) by the (morphologically free) particle *je*; 2) by the use of the interrogative roots *-ni* and *-pi* which take various prefixes to form question words functioning as adverbs (of time - *lini*, place - *wapi*, and manner - *vipi*), as pronouns (*nani* - who, *nini* - what, *-pi* - which one, *-ngapi* - how many) or as adjectives (*gani* - what sort, *-pi* - which, *-ngapi* - how many); 3) by the use of *-je* (manner), *-ni* (neuter) or *-pi* (place) as bound morphemes suffixed to the verb, (*-ni* and *-pi* are rarely found, however). Below, I shall look at the free particle *je* and the bound morpheme *-je*, as used in the first and third question forms respectively.

König (1991:13), in his discussion of what he called 'focus particles', observed that,

"In languages in which yes-no interrogatives are distinguished from declaratives not through word order, but through the addition of certain particles, these particles are often added to the focused constituent."

Questions in Kiswahili may be optionally indicated by the occurrence of *je*. Rather than being added to a simple constituent (a verb or noun in isolation, say), *je* occurs either initially, where it forms a separate tone group (Maw & Kelly 1975:58), or finally in a clause, and has scope over an utterance expressing a proposition. What is

questioned is the entire proposition expressed (however explicit or implicit the utterance) rather than any particular part of it. The enclitic *-je*, on the other hand, is added to a verb phrase; the focused constituent is therefore the verb phrase alone, but the information requested is not confirmation or disconfirmation, but rather further information about the action associated with that verb phrase. This latter form is more akin to a *wh*-question than a yes-no question. I shall describe first the free particle *je* and then the bound morpheme *-je*, and evaluate two hypotheses suggested by the data.

*Je* resembles the particle *ré* in Sissala, a Niger-Congo, Gur (Voltaic) language spoken in Burkina Faso and Ghana (Blass 1990), in so far as questions in Sissala may be optionally marked with *ré* (Blass 1990:117). However, unlike *je*, *ré* can occur in both questions and in answers, frequently marks figurative speech (i.e. irony), beliefs and desires, and occurs in embedded 'perceptual' clauses (with verbs meaning *see*, *hear*, etc.). Nonetheless, the underlying function of *ré* is similar to that of *je*, even though the range of interpretations derivable from these two particles is not coextensive; Blass characterises the underlying function of *ré* as follows:

"My proposal is that *ré*, and other 'hearsay' particles in other languages, have the function of indicating to the hearer that the utterance which contains them is interpretively used. In other words, *ré* is an explicit linguistic indicator of interpretive use."

(Blass 1990:104)

My proposal is that the underlying function of *je*, like that of *ré*, is as a linguistic indicator of interpretive use. By the interpretive use of an expression (as opposed to its descriptive use) Blass means the use of a propositional form "to represent not itself, but some other propositional form it more or less closely resembles" (S&W 1986:264 n.25), such as incredulity or doubt as to the truth of the proposition expressed. The distinction is similar to the use/mention distinction employed in Sperber & Wilson's (1981) analysis of irony, except that there need be no full linguistic or logical identity between the original propositional form and what it represents where interpretive use is concerned. In the case of *je*, interpretive use is restricted to the interrogative function (although see examples (2.44) and (2.45) and discussion), but there is considerable variation in the kinds of question expressed by utterances containing *je*. The following examples serve to illustrate this:

- (2.39) "Je, Najum, utakwenda sokoni?"  
*Je, Najum, you will go to the market?*  
 "Najum, are you going to the market?"

(MWK p.14)

- (2.40) "Unatakiwa kwa Mkuu," mmoja wao alimweleza.  
*"You're wanted at the Governor's," one of them explained to him.*  
 "Je, kwema?"  
*Je, good? "Is that a good place?"* (Uk p.19)
- (2.41) Tunajikinga na Ukimwi. Wewe je? (Tanzanian T-shirt slogan)  
*We struggle with Aids. You je?*  
*We are fighting Aids. What about you?*

*Je* in (2.39) indicates that the following proposition, 'You will go to the market', is to be used interpretively, giving rise to an interrogative interpretation. In (2.40) *Je* indicates that what is grammatically an adjective (consisting of an agreement marker *ku-* in concord either with the infinitive of the preceding verb or with the locative *kwa*, and an adjectival stem *-ema*, 'good'), is being questioned. Finally, in (2.41) *Je* questions the emphatic pronoun 'you', which the addressee is expected to complete into 'you are also fighting Aids'. *Je* can also occur alone in an utterance; in such cases its interpretation depends, as ever, on the interpretation context, and can convey ideas which Johnson (1939) glosses as "How? Well? What now? Answer me! Tell me! etc."

The bound form, *-je*, is open to a similarly wide range of interpretations, depending on the nature of the verb to which it is bound and the interpretation context. Often, questions indicated by the *-je* enclitic can be paraphrased more explicitly by the use of one of the question words, e.g. *vipi?* (how?) could have been used in examples (2.42) and (2.43) second sentence, or *nini?* (what?) could have been used in (2.43) first sentence. (Instances of *-je* are underlined):

- (2.42) "Unaionaje hali yako sasa? Umeumia sana?"  
*You see-je your condition now? You are hurt very much?*  
*"How do you feel now? Are you badly hurt?"* (MWK p.25)
- (2.43) "Nitafanyaje kuitisha wito wa kujitawala na kuitikiwa na Wanyika?  
 Itakuwaje iwapo nitashindwa kuwashawishi kujiunga pamoja na kudai utawala wa haki na kwa kweli?"  
*"What shall I do (literally: I shall do -je) to bring about the calling for self rule and to be called (to leadership) by the Wanyika? What will happen (literally: It shall be-je) if I fail to persuade them to join together and truly claim their rightful (self) rule?"* (Uk p.12-13)

In the light of the above data, I shall propose two hypotheses, one specific to Kiswahili interrogative forms and the other more general. They are:

1) The free and bound forms *je* and *-je*, described above, are exponents of a single particle, JE, which has the unitary and underdetermined semantic form to be described below.

2) If an interpretation context is already highly constrained, fewer or less explicit procedural indicators are required to fulfil the criterion of consistency with the principle of relevance, that "Every act of ostensive communication communicates the presumption of its own optimal relevance." (S&W 1986:158). For example, *-je* is less explicit than *nini*, since *nini* could have been used in place of *-je* in example (2.43) to mean 'what' but not in (2.42) to mean 'how'.

JE is semantically underdetermined, and the range of potential interpretations which can be assigned to utterances containing it depends on its scope<sup>25</sup> and the interpretation context. Wilson & Sperber (1993a:22) suggested that illocutionary force indicators, such as *je* and *-je*, encode procedural constraints on the inferential construction of higher-level explicatures. Semantically, JE is *an instruction to the addressee to treat whatever is within its scope as being used interpretively by the speaker for the purpose of seeking more information about it*. This paraphrase of the semantics of JE is deliberately vague because JE itself is significantly semantically underdetermined. The term 'whatever is within its scope' was used rather than 'the proposition within its scope' because it is often not a proposition which is interpretively used, but something less determinate. A clause (representing a full proposition) is interpretively used in (2.39), however, elsewhere it is an adjective (2.40), an emphatic pronoun (2.41), or a verb (2.42) and (2.43) that is interpretively used. When *Je?* occurs alone in an utterance it could be said to have the interpretation context as a whole in its scope. In order to arrive at the correct interpretation of the speaker's utterance the addressee must perform two inferential tasks: she must enrich whatever is in the scope of JE, or in cases where there is nothing to enrich, must recover the intended mutually manifest assumption; and she must enrich JE to identify the kind of information being sought by the speaker.

The second hypothesis, that if an interpretation context is already highly constrained, fewer or less explicit procedural indicators are required to fulfil the criterion of consistency with the principle of relevance, is suggested by the facts, a) that JE is optional, and b) that more precise question words also exist in Kiswahili. As already mentioned, a yes-no question in Kiswahili is generally indicated by intonation alone, and other question forms involving JE often include - or have more explicit

---

<sup>25</sup> By whatever is within the scope of JE, I mean the focused constituent, in the sense of König (1991) above. At the level of linguistic semantic representation (the output of phase one) JE has scope over conceptual representations which may be either complete or incomplete logical forms.

paraphrases which include - the question words *nini?* (what?), *vipi?* (how?) etc. Only when the interpretation context is sufficiently constrained (that is, when the most relevant set of mutually manifest assumptions is obvious) will fewer procedural indicators (e.g. intonation alone) or less explicit procedural indicators (e.g. *-je* as opposed to a more explicit question word) be used. For example in (2.42) repeated below, the first question "How do you feel?" is introduced by *-je*, but the second question "Are you badly hurt?" does not contain *je*, as a question sequence has already begun.

- (2.42) "Unaionaje hali yako sasa? Umeumia sana?"  
*You see-je your condition now? You are hurt very much?*  
*"How do you feel now? Are you badly hurt?"*

(MWK p.25)

There are two sides to the notion of optimal relevance: adequate contextual effects, and minimal processing effort (see §1.4.3). The above hypothesis concerns the reduction of processing effort; when an interpretation context is already highly constrained, explicit procedural indicators like JE ought to be superfluous. However, JE does occasionally occur in utterances which are already highly constrained, such as the following examples:

- (2.44) Je, mzee hali yake vipi siku hizi?  
*Je, old person his/her state how these days?*  
*So/By the way, how is grandfather (etc.) getting on these days?*
- (2.45) "Je, nini khabari Bin Malik? Sikupata kukuona hata siku moja ofisini kwetu; je, kumezidi nini tena leo?"  
*Je, what news Bin Malik? I didn't get to see you even one day in our office; je, what has increased again today?*  
*So, what news (do you have) Bin Malik? I haven't been able to find you in our office even once; so what's come up again this time?*

(KMF p.55)

Because of the occurrences of the question words *vipi* in (2.44) and *nini* in (2.45) *je* is superfluous as an interrogative indicator, since no other interpretation of propositional attitude other than that of a question is possible; hence *je* is not necessary to reduce processing effort. The principle of relevance suggests that the addressee will seek additional contextual effects to compensate for the additional processing effort involved in processing *je*. In both cases, the additional contextual effect achieved by the addition of *je* to the questions is stylistic. In (2.44) *je* is glossed as *So/By the way* since an utterance of (2.44) would be appropriate on a first mention of *mzee* by the speaker; in such circumstances *je* acts as a topic change indicator. *Je* in this example

is not totally superfluous from a cognitive point of view, since the other interrogative marker, the question word *vipi*, doesn't occur until late in the sentence. This is in contrast to the question word *nini* in (2.45) which occurs sentence initially in the first instance. The presence of *je* in (2.45) indicates that the questions 'what news Bin Malik?' and 'what's come up again this time?' are being mentioned. Yet they are not mentioned to request further information (they are already unambiguously questions), but rather to express another interpretive function; the most relevant interpretation with respect to the utterance interpretation context, in this case the passage from which the example was taken, is an ironic interpretation. This is not surprising if we consider the aforementioned similarities with the interpretive use indicator *ré* in Sissala which is also used to mark irony.

It is not necessary to revise the characterisation of the semantics of JE as an essentially interrogative marker in the light of (2.44) and (2.45). JE merely encodes that the question has been mentioned rather than used, and (since an interrogative interpretation is blocked by the fact that this is already the only possible interpretation of the mentioned constituent), a further interpretation must be sought. Irony, however, is not encoded by JE; it is simply the most relevant interpretation in this particular context.

### *Summary*

I have defined procedural information at phase three as information which constrains the computation of contextual effects without directly contributing to them. Contextual effects arise out of the contextualisation of new information {P} in existing assumptions {C}. The function of *jel-je* in a given utterance is to indicate that the sentence expressing the proposition which constitutes {P} has been interpretively used. Unless other more explicit question words (*vipi*, *nini*, etc.) have been used to form a question, *jel-je* indicates that a proposition or incomplete semantic representation has been interpretively used in order to request more information about it (confirmation or disconfirmation of its truth, information concerning how, what, etc.). When explicit question words do occur, *jel-je* still indicates that whatever is within its scope has been interpretively used, but the addressee is free to look beyond an interrogative interpretation and, in line with the principle of relevance, recover additional contextual effects, such as indication of a change of topic or irony.



## 2.3 Conclusions

### 2.3.0 Introduction

We have seen that exponents of procedural encoding at phase three can take a variety of linguistic forms. The discourse connectives *so* and *well* are discrete lexical items occurring outside of clause structure; the Kiswahili interrogative marker JE can be realised either as the morphologically free form *je* preceding or following a clause or as the morphologically bound verbal suffix *-je*. Procedural information may be encoded syntactically in the form of mood and information packaging, and may also be specified phonetically through intonation. The distinction between procedural encoding constraining implicatures and procedural encoding constraining higher-level explicatures cross-cuts these linguistic forms: constraints on implicatures are encoded by *so* and *well* and by information packaging; constraints on higher-level explicatures are constrained by intonation, syntactic mood and by *je/-je*.

In this final section, I shall provide a theoretical definition of procedural encoding at phase three and state precisely how the examples discussed above correspond to this definition. I shall then propose criteria by which exponents of the different types of procedural encoding at phase three can be distinguished descriptively from each other and from conceptual encoding.

### 2.3.1 Procedural encoding at phase three: a theoretical characterisation

The output of phase three of utterance interpretation is a contextual effect or effects; hence, procedural encoding at phase three constrains the computation of contextual effects. Since contextual effects result from the contextualisation of new information {P} in existing assumptions {C} (the utterance interpretation context) there are various ways in which the computation of contextual effects could, theoretically at least, be constrained. To begin with, {C} is constructed rather than given; although any {P} is almost certain to achieve some contextual effects in combination with all the addressee's currently held assumptions, it would do so at the cost of a large amount of processing effort. For this reason, consistency with the principle of relevance (§1.4.3) ensures that if {P} achieves adequate contextual effects in combination with those assumptions currently most accessible to the addressee, the interpretation process comes to an automatic halt, thus minimising processing effort.

Any assistance the speaker can give the addressee in selecting the subset of assumptions {C} in which a contextualisation of {P} will achieve adequate contextual effects will reduce the addressee's processing effort. The discourse connective *well*

performs such a function by indicating that the assumptions which the speaker assumes to be currently most accessible to the addressee (that is, those assumptions from which the speaker assumes the addressee will construct {C}) are inappropriate for a successful contextualisation of {P}. *Well* constrains the construction of {C} and hence reduces the processing effort required by the addressee to derive contextual effects. Information packaging also aids the addressee in the construction of {C}. Word order and the choice of, for example, definite descriptions as opposed to pronouns, all assist the addressee in the construction of {C}. The new information which constitutes {P} is also specified by syntactic information packaging through the use of topicalisation, clefts, pseudo-clefts, etc. and by intonation.

Having constructed an appropriate utterance interpretation context {C} and identified the new information {P} which the speaker intends to communicate, the addressee still has to relate {P} to {C} in such a way that adequate contextual effects result. The use of the discourse connective *so* assists the addressee in this process by instructing the hearer to establish an inferential connection between {P} and {C}. Markers of interpretive use, such as syntactic mood and the Kiswahili interrogative particle *je/je*, also contribute to this process by constraining higher-level explicatures in the form of the propositional attitudes associated with proposition(s) in {P}.

In §2.1.2 I proposed that both conceptual and procedural information are represented at the level of linguistic semantic representation (the output of phase one of utterance interpretation). An expression which encodes conceptual information contains in its logical selection frame information about itself, whilst an expression which encodes procedural information contains information about the interpretation of whatever is within its scope. This may be either a fully propositional conceptual representation (that is, a conceptual representation of a determinate state of affairs which provides the input to phase three of utterance interpretation) or a sub-propositional conceptual representation (constituting the input to phase two of utterance interpretation). In the terminology of Wilson & Sperber (1993a:2), procedural encoding at phase three 'constrains the manipulation' of fully propositional conceptual representations.

Defining procedural encoding at phase three in terms of expressions which constrain the manipulation of fully propositional conceptual representations does not of itself provide us with adequate heuristic criteria by which to identify exponents of procedural encoding at phase three. If the theoretical distinction is to be of use in descriptive linguistics we must be able to identify linguistic expressions which encode procedural information constraining phase three of utterance interpretation; specifically we require criteria by which to distinguish procedural encoding from

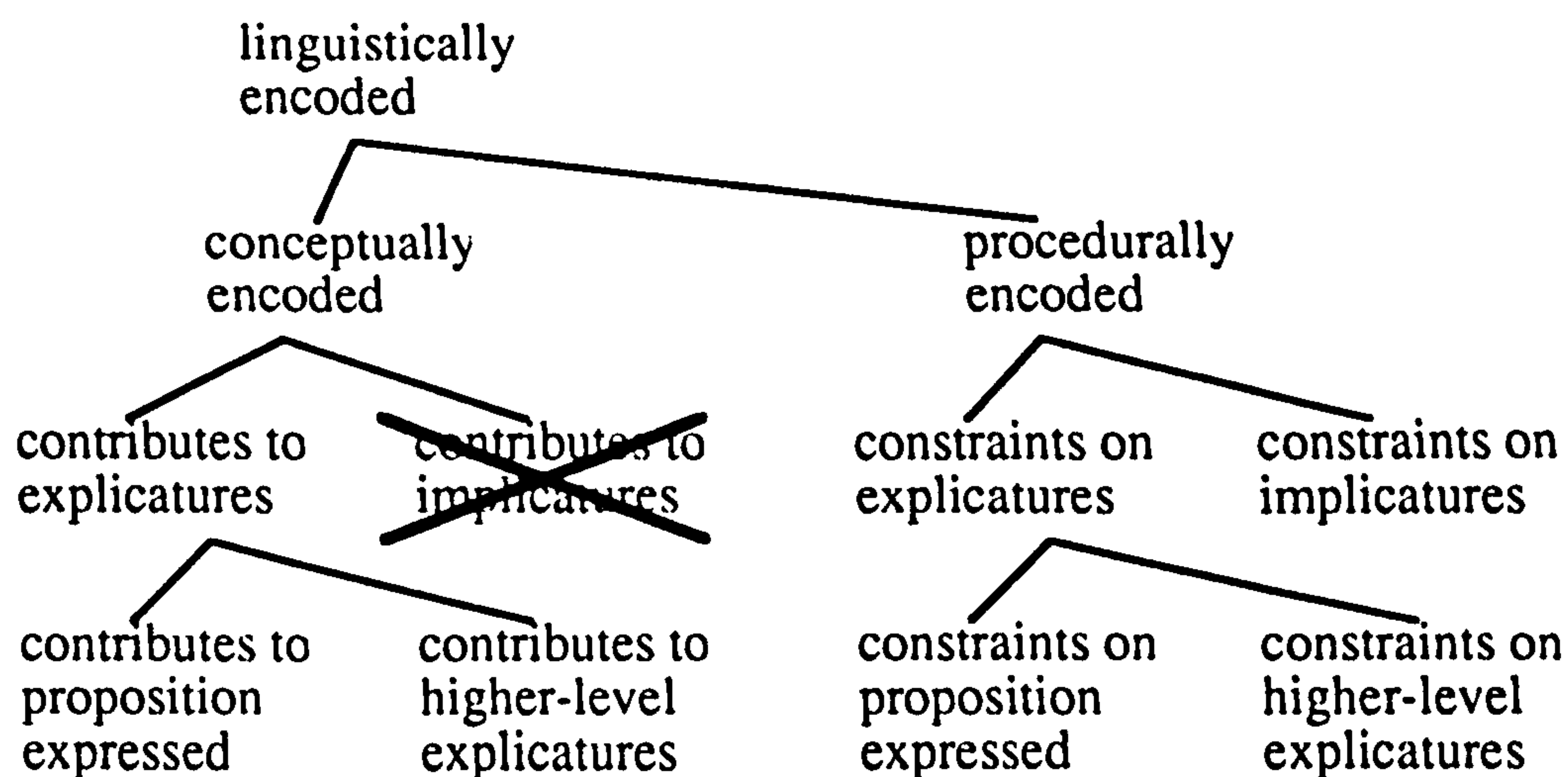
conceptual encoding, and constraints on implicatures from constraints on higher-level explicatures.

### 2.3.2 Procedural encoding at phase three: a descriptive characterisation

In this section, I shall propose criteria by which exponents of procedural encoding at phase three of utterance interpretation may be distinguished from their conceptually encoded counterparts.

Figure 7 below illustrates the types of linguistically encoded information which must be distinguished descriptively as well as in the theory:

Figure 7: Possible types of linguistically encoded information.



The subcategory of conceptually encoded information which contributes to implicatures has been ruled out following my argument in §2.1.1 against the existence of exponents of this subcategory. Therefore, the subcategory of linguistically encoded constraints on implicatures (of which *so*, *well* and information packaging are exponents) has no direct conceptual counterpart.

In contrast to exponents of procedural encoding constraining higher-level explicatures, such as non-declarative word order in English and the *je* interrogative marker in Kiswahili, Wilson & Sperber (1993a:16-19) posit a class of conceptually encoded expressions which contribute to higher-level explicatures; these include illocutionary adverbials such as *seriously*, *frankly* and *confidentially*, and attitudinal adverbials such as *unfortunately*:

- (2.50) a. Seriously, I can't help you.  
 b. Frankly, I can't help you.  
 c. Confidentially, I can't help you.  
 d. Unfortunately, I can't help you. [W&S 1993a (18)]

Whereas exponents of procedural encoding constraining higher level explicatures contribute to the truth conditions of neither the associated proposition nor the propositional attitude, the higher-level explicatures to which illocutionary and attitudinal adverbials contribute are "conceptual representations, capable of entailing and contradicting each other and representing determinate states of affairs. Though true or false in their own right, they do not generally contribute to the truth conditions of their associated utterances." (Wilson & Sperber 1993a:16) These conceptual expressions are preserved in indirect reported speech (unlike interrogative word order in most dialects of English and *je/-je* in Kiswahili); illocutionary adverbials appear as their "synonymous manner-adverbial counterparts" (W&S 1993a:17) and contribute not to higher-level explicatures but to the proposition expressed:

- (2.51) a. Mary told Peter seriously that she couldn't help him.  
 b. Mary said frankly to Peter that she couldn't help him.  
 c. Mary informed Peter confidentially that she couldn't help him.  
 [W&S 1993a (19)]

whilst attitudinal adverbials continue to contribute to higher-level explicatures:

- (2.51) d. Mary told Peter that, unfortunately, she couldn't help him.

They are also semantically complex in comparison to procedural illocutionary force indicators. The distinction between procedural constraints on higher-level explicatures and conceptual contributions to higher-level explicatures in English with respect to semantic complexity is illustrated in (2.52):

- (2.52) a. Seriously, are you leaving? [W&S 1993a (20)]  
 b. She asked me seriously whether I was leaving.  
 c. She asked me to tell her seriously whether I was leaving.

Wilson & Sperber (1993a:18) point out that if (2.52a) is interpreted as a request to tell, *seriously* is ambiguous in that it can modify either the requesting or the telling; these alternative interpretations are reflected in (2.52b) and (2.52c) respectively. Wilson & Sperber (1993a:18) conclude: "It is not obvious how this ambiguity could

be handled in procedural terms." The theoretical characterisation of procedural encoding at phase three provided in §2.3.1 explains why it is not obvious how this ambiguity could be handled in procedural terms. Given that procedural encoding gives rise to logical selection frames providing information about the manipulation of conceptual representations within their scope, it is necessary for the extent of that scope to be specified (either within the logical selection frame itself or by associated syntactic or intonational information). Hence, all procedural encoding must specify, (a) information about the manipulation of any conceptual representation(s) within its scope, and (b) the precise extent of that scope, including the nature of the conceptual representation about which information is provided. The illocutionary adverbials fail to specify (b) and hence any attempt to characterise them in procedural terms will encounter problems.

### 2.3.3 Summary

In this chapter, the notion of a distinction between conceptual and procedural encoding was introduced, and it was suggested that this distinction must, if linguistically encoded, be represented within the language of thought. Conceptual information was viewed as contributing to the construction of conceptual representations, whilst procedural information was viewed as constraining the manipulation of conceptual representations. The exponents of procedural encoding which I discussed in §2.2 aid the addressee in the identification of implicatures and higher-level explicatures. This has the effect of reducing inferential processing effort at phase three of utterance interpretation and hence contributes to relevance (the smaller the amount of processing effort required to process an utterance the greater its relevance). In chapter 3 I shall defend the theoretical characterisation of procedural information as constraining the manipulation of conceptual representations and consider in particular the manipulation of sub-propositional conceptual representations to constrain the proposition expressed.

### 3 PROCEDURAL ENCODING AT PHASE TWO

In this thesis, utterance interpretation has been viewed as a three-phase process: a modular decoding phase results in a semantically underdetermined logical form which is inferentially enriched to give rise to a fully propositional form, this is then contextualised in a subset of the addressee's existing assumptions resulting in contextual effects. In §2.2 I characterised a variety of lexical and syntactic constructions as exponents of procedural encoding at phase three, and demonstrated how these constrain the interpretation of fully propositional conceptual representations by constraining the inferential contextualisation process. In this chapter I shall discuss procedural encoding at phase two, which constrains the inferential enrichment of sub-propositional conceptual representations. In §2.1.2 we saw that, in principle, it is possible for linguistically encoded constraints on interpretation to constrain the interpretation of elements within a logical form. Given that logical forms are blueprints for propositions, individual elements within logical forms must be sub-propositional, hence any procedural encoding which ranges over the interpretation of such elements manipulates sub-propositional conceptual representations. The manipulation of sub-propositional conceptual representations constrains the construction of propositions, which is the output of what I have termed phase two of utterance interpretation; I shall therefore speak of such linguistically encoded constraints on interpretation as procedural encoding at phase two (in contrast to procedural encoding at phase three).

Thus far, procedural information at phase two has simply been posited without reference to the kinds of expressions which might encode it. Wilson & Sperber (1993a) have suggested that (English) personal pronouns be viewed as exponents of procedural encoding at what I have termed phase two. This claim will be addressed and it will be shown that, in the absence of a satisfactory truth-conditional account of the semantics of pronouns, a procedural account provides a viable alternative. This procedural analysis of English personal pronouns will be applied to the analysis of pronouns in Kiswahili and extended to encompass tense, aspect and (in chapter 4) modality, which have also defied adequate characterisation in directly truth-conditional terms.

The main purpose of this chapter is to demonstrate that a procedural account of tense and aspect (derived from a procedural analysis of pronouns) is viable, and satisfactorily accounts for the use of tense and aspect markers in discourse. I shall demonstrate for each of pronouns, tense and aspect, that these cannot be directly truth-conditionally interpreted, since they significantly underdetermine the

propositional forms to which they contribute. The precise identification of an intended propositional form is largely dependent on pragmatic considerations, and I will suggest that the function of pronouns, and tense and aspect markers is to constrain the inferential (pragmatic) processes involved in the identification of propositional form.

The detailed outline of this chapter is as follows: §3.1 provides a characterisation of pronouns as exponents of procedural information, incorporating Ariel's Accessibility Hierarchy and extending the procedural analysis of pronouns to Kiswahili. Whilst pronouns constrain the interpretation process with respect to discourse entities, tense aspect and modality markers are involved in the interpretation of situations and events. §3.2 provides a theoretical introduction to how situations and events are represented through the interaction of sub-propositional conceptual representations, procedural information, and the utterance interpretation context. In §3.3, the role of tense and aspect markers in the determination of temporal reference will be discussed, and I will propose that these encode procedural information constraining the interpretation of sub-propositional conceptual representations of situations and events. In §3.4 I shall summarise the findings of the previous three sections prior to providing a procedural analysis of the modal auxiliaries *may*, *can*, *must* and *should* in chapter 4.

## **3.1   Pronouns**

### **3.1.0   Introduction**

The use of pronouns is linked to the more general problems of reference assignment and anaphora resolution. There are numerous theories of anaphora resolution, but it is recognised that this cannot be reduced to syntactic considerations alone (GB simply states that pronouns are free within their governing category) or resolved apart from inferential (pragmatic) considerations.<sup>26</sup> In addition, pronouns are variables which are not directly truth-functionally interpretable; truth-conditions are properties of propositions, and pronouns only contribute to propositions once the inferential process of reference assignment has taken place (unless they remain within the scope of quantifiers as is the case with 'donkey sentences').

---

<sup>26</sup>The concept of anaphora and the question of anaphoric reference assignment (including how to account for reference assignment in regard to pronominal anaphors lacking explicit linguistic antecedents) have been addressed within the relevance theoretic framework by Kempson (1988b, 1988c, 1994a, 1994b) and Reboul (1994; and references therein, in particular to the work of Kleiber). The relation between syntax and pragmatics as this affects the binding of pronouns is discussed in Reinhart (1983) and, from a relevance theoretic perspective, Foster-Cohen (1994).

That reference assignment is a largely inferential process, is evident from Chinese; although it is a pro-drop language, verbs in Chinese do not mark agreement morphologically, so in the interpretation of an utterance of a sentence with a zero-pronoun (the majority of third person pronouns are unrealised), reference assignment is purely pragmatic (Li & Thompson 1979). This is also illustrated in the following English examples, the interpretation of which depends on the addressee's knowledge of the typical behaviour of people who are happy or who see others in distress:

- (3.1) a. Jane<sub>i</sub> hugged Mary<sub>j</sub> because she<sub>i</sub> was happy.  
b. Jane<sub>i</sub> hugged Mary<sub>j</sub> because she<sub>j</sub> was crying.

However, there are restrictions on reference assignment to pronouns which appear to be conventionalised, hence the unstressed *she* in (3.2) is coindexed with Jane rather than with Mary or some other female not mentioned in the discourse:

- (3.2) Jane<sub>i</sub> hugged Mary<sub>j</sub> and then she<sub>i/\*j/\*k</sub> hugged Harry.

In the absence of a directly truth-conditional analysis of pronouns, I suggest that a satisfactory characterisation of the semantics of pronouns can be formulated only if the conceptual/procedural distinction is taken into account. Below, I shall outline an account of pronouns which analyses restrictions on reference assignment such as those in (3.2) as resulting from procedural information encoded by pronouns. Since reference assignment contributes to the propositional form recovered, pronouns will be treated as exponents of procedural encoding at phase two of utterance interpretation.

Wilson & Sperber's (1993a) procedural account of pronouns (which forms part of the typology of information types introduced in §2.1.1) will be assessed and integrated with Ariel's Accessibility Hierarchy (Ariel 1988, 1990, 1991, 1994).

### 3.1.1 A procedural account of pronouns

The idea that pronouns and other types of referring expression might encode procedures predates W&S (1993a). For example, Hawkins (1978:17, cited in Matsui 1995:10) suggested that "the use of the definite article acts as an instruction to the hearer to locate the referent of the definite NP" by searching for it in "the appropriate, pragmatically identifiable, set" (ibid.), and Ariel, whose theory of Accessibility I shall shortly discuss, proposed that,



"instead of claiming that an expression of type *x* is processed in a certain way... we view the processing procedure associated with each form as its inherent definition. In other words, referring expressions are no more than guidelines for retrievals."

(Ariel 1988:68)

A similar account is suggested from a philosophical perspective by Kaplan (1989:523), who claimed that pronouns encode rules "which *determine* the content (the propositional constituent) for a particular occurrence of an indexical. But they are not *part* of the content (they constitute no part of the propositional constituent)." Wilson & Sperber (1993a) adopt Kaplan's argument (without buying into any theory of Direct Reference) and interpret it as a claim that pronouns encode procedural information. The justification for this characterisation of pronouns rests on the following argument (given in full):

"the claim that 'I' means 'the speaker' has different consequences depending on whether it is conceptually or procedurally understood.

Suppose that David Kaplan says (3.3):

(3.3) I do not exist.

Then if 'I' is treated as encoding the *concept* the speaker, (3.3) will express the proposition in (3.4):

(3.4) The speaker of (3.3) does not exist.

But if 'I' is treated merely as encoding an instruction to identify its referent by first identifying the speaker, then (3.3) will express the proposition in (3.5):

(3.5) David Kaplan does not exist.

These two propositions differ in their truth conditions. (3.5) is true in any state of affairs in which David Kaplan does not exist. (3.4) is true in any state of affairs in which (3.3) is uttered and its speaker does not exist. Since such a state of affairs is impossible, if (3.3) expressed the proposition in (3.4), it would be necessarily false. Kaplan argues that although (3.3) is false whenever it is uttered, it is not *necessarily* false. The proposition it expresses is true in any state of affairs in which David Kaplan does not exist. In other words, (3.3) must be understood as expressing (3.5) and not (3.4)."

(Wilson & Sperber 1993a:20)

Kaplan is correct in claiming that (3.3) is not necessarily false, but the argument cited simply stipulates this, but does not prove that (3.3) is not necessarily false. Kaplan's argument can be made more explicit as follows. Suppose that I (Steve Nicolle) go to the local tax office and find that there is no record of my National Insurance number. If, on recounting this story to a friend, I utter (3.3), this will be felicitous if understood as (3.6):

- (3.3) I do not exist.
- (3.6) The individual named Steve Nicolle does not exist as far as the tax office is concerned.

For this interpretation to be recovered, the pronoun *I* must encode procedural information which licenses the addressee not just to identify the referent of *I* as 'the speaker', but to go beyond this and identify the speaker as a particular individual (i.e. the individual named Steve Nicolle). In other words, the pronoun *I* tells the addressee "to identify its referent by first identifying the speaker." (*ibid.*) If *I* encoded the concept **the speaker**, (3.3) would express the proposition expressed in (3.4), which would be infelicitous even in the scenario just described.

However, the above characterisation of the semantics of pronouns fails to address the following question: given, as argued in §2.1.2, that procedural encoding at phase two (of which pronouns are exponents) encodes information about the manipulation of sub-propositional conceptual representations, what kind of sub-propositional conceptual representations are being manipulated by pronouns?

It is tempting to suggest that the sub-propositional conceptual representation whose interpretation a pronoun manipulates is a mental representation of the intended referent. However this position is untenable. The intended referent is indeed represented in the mind of the addressee, but as an element of the utterance interpretation context. Pronouns simply guide the addressee to the intended referent; once this has been established the role of the pronoun is complete.

So what is the sub-propositional conceptual representation whose interpretation is constrained by a pronoun, and where does it originate? One potential solution is to deny that there is any conceptual representation to be manipulated; this is the approach adopted in W&S (1993a). Whilst Wilson & Sperber (1993a:2) begin by suggesting that there might be a category of procedural expressions which encode information about the manipulation of conceptual representations, they conclude by proposing that these procedural expressions "constrain the inferential phase of comprehension by reducing the hypothesis space to be searched in arriving at the intended interpretation" (W&S 1993a:21). This latter characterisation of procedural encoding, in which no mention is made of the manipulation of conceptual representations is intended to supersede that in which the manipulation of conceptual representations is invoked (Deirdre Wilson p.c. 28/10/95).

However, there are good reasons for preferring the model in which conceptual representations are manipulated over the model proposed in (W&S 1993a:21) in

which no mention is made of conceptual representations. Firstly, an account of pronouns in which no mention is made of conceptual representations predicts ill-formed and incomplete logical forms. Kokolakis (1994:15) suggests (in line with W&S 1993a:10) that (3.7) has a logical form of the type (3.8), which is well formed:

- (3.7) Peter told Mary that the weather was nice.  
(3.8) x told y at  $t_1$  that the weather was nice at  $t_2$

If pronouns encode only procedural information, however, an utterance of (3.9) would have a logical form of the type (3.10) which is ill-formed and cannot therefore act as the input to logical inference rules or enter into relations of entailment or contradiction:

- (3.9) He told her that the weather was nice.  
(3.10) [procedure] told [procedure] at  $t_1$  that the weather was nice at  $t_2$

Now, Kokolakis (1994) does not state what logical forms consist of, so his argument that (3.10) is not a well-formed logical form is incomplete. However, in §2.1.2 I argued, in line with Groefsema (1992), that any information represented in a logical form must be capable of being represented in the language of thought. On this view, procedural information consists of semantic representations providing information, not about the semantic content of the expressions which encode them, but about conceptual representations within the scope of these expressions. Thus, (3.10) can be shown to be ill-formed because it is not stipulated what the conceptual representations are whose interpretations the two [procedure]s constrain.

Secondly, in §1.2.1 I observed that Sperber & Wilson (1986:66, 75) and Wilson & Sperber (1986a) claim that an account of utterance interpretation offers an insight into other, more complex functions of the central systems, such as scientific theorising (and presumably of less specialised but more general problem solving activities) hence, the inferential phases of communication are manifestations of more general inferential abilities which are employed not only in utterance interpretation but in other problem solving activities. In addition to this, Handley & Buck (1995) demonstrated experimentally that people's general inferential problem solving abilities function through the manipulation of conceptual representations. We would expect, therefore, that the manipulation of conceptual representations would feature throughout the utterance interpretation process also.

Two points can be drawn from the above discussion: first, that pronouns encode procedural information licensing an addressee to identify a particular individual as the intended referent; and second, that there must be some class of conceptual representations whose interpretations pronouns constrain. I have reviewed some of the evidence in favour of retaining the manipulation of conceptual representations as integral to my characterisation of procedural encoding, and have rejected as a possible candidate a mental representation of the intended referent, but I have still not answered the question asked earlier about the nature and origin of the sub-propositional conceptual representation which is manipulated by a pronoun.

Recall that in §2.1.2 it was suggested that, in principle, a single expression might encode both conceptual and procedural information. Could it be that pronouns encode both information types: conceptual, providing the conceptual representation to be manipulated, and procedural, providing constraints on the manipulation of this conceptual representation? Pronouns in English encode information concerning person, number, gender and case. However only person is encoded by all the personal pronouns in English: *you* specifies that an instance of second person reference is at stake, but fails to distinguish number, gender or case (this latter being specified structurally). Given that the procedural information encoded by *I* takes the form of an instruction to identify its referent by first identifying the speaker (i.e. the first person), information concerning person must be procedural. Since *you* encodes no more information than that concerning person, it must be an exponent of procedural encoding alone, and so this pronoun at least encodes no conceptual information.

Instead, I propose that the conceptual representations which are manipulated in line with the procedural information encoded by pronouns are projected from the logical entries of predicates, which specify what the predicate needs to combine with to yield a well-formed logical form (see §1.3.1). The lexical entries of concepts also specify, among other things, syntactic co-occurrence restrictions in the form of strict subcategorization frames, but, as Groefsema (1995b) demonstrates through a consideration of implicit arguments in English, it is only the logical selection restrictions of a predicate which determine its arguments.<sup>27</sup> For example, the logical entry of TELL specifies that TELL is a predicate which requires an agent (to do the telling), an instance of a patient (what is told) and an instance of a recipient (the person who is told). The patient can be left implicit if it is highly accessible in the

---

<sup>27</sup> Whether an argument can be left implicit depends on whether the logical entry of a predicate plus the utterance interpretation context gives rise to an interpretation in accordance with the principle of relevance. For discussion in relevance theoretic terms of the role of verb subcategorization in utterance interpretation see also Groefsema (1992) especially chapter 3.

utterance interpretation context, such as during a discussion of some piece of news ('Mary told me \_'). Thus, *told* in (3.9) (repeated below) projects an extremely semantically underdetermined conceptual representation of an agent into the subject position of the clause and a patient and/or recipient into the direct and/or indirect object position. The identification of the intended discourse referents is constrained by the procedural information encoded by the pronouns in subject and object positions.

(3.9) He told her that the weather was nice.

(3.11) [Agent] told [recipient] at  $t_1$  that the weather was nice at  $t_2$

On this account, the absence of an overt pronoun in subject position in *pro*-drop languages causes no problem, since subjects in logical forms are projections of the predicate; this account is in line with the emergent Minimalist Programme (Chomsky 1991, 1993, 1995) in which lexical items and their morphology are inserted 'ready made' into syntactic structure, where inflectional nodes, rather than adding inflections simply check them.

In addition to procedural information concerning person, case, etc. which differs between pronouns, all members of the set of pronouns (in a given language) encode procedural information relating to the relative accessibility of intended discourse referents. This takes the form of an instruction to find a highly accessible discourse entity within the utterance interpretation context, which is compatible with both the specific procedural information encoded by a particular pronoun and the conceptual representation projected by the logical entry of the relevant predicate.<sup>28</sup> This information is subsumed by the particular procedural information encoded by first and second person pronouns (an instruction to identify the intended referent by first identifying the speaker or addressee) since speakers and addressees are always highly accessible within an utterance interpretation context. In the case of third person pronouns, however, the general procedural information relating to accessibility distinguishes the referents of pronouns from those of descriptions, names, demonstratives, emphatic pronouns, etc. In §3.1.2 below, I shall discuss the notion of accessibility.

---

<sup>28</sup> Such a characterisation of pronouns is in line with the experimental evidence, such as that relating to the 'repeated name penalty', cited in the discussion of the status of unstressed pronouns in information packaging (§2.2.2), and evidence from cross-modal priming experiments reported in Nicol & Swinney (1989) (cited in Groefsema 1995b:155) that possible referents in a context are immediately activated when a pronoun is processed.

### 3.1.2 Accessibility

The notion of accessibility is taken from Accessibility theory (Ariel 1988, 1990, 1991, 1994), according to which, referring expressions, including pronouns, encode the relative degree of accessibility of mental representations.<sup>29</sup> The accessibility, or salience, of mental representations of discourse entities is determined by a combination of various factors; Matsui (1995:45-60) reviews a number of studies which address the effect of order of mention, recency of mention, syntactic position, thematic role, and the semantics of the main verb on accessibility; topicality, the number of possible alternative referents and physical saliency may also play a role.

Matsui (1995:53-4) also mentions experimental evidence from Sanford *et al* (1988) that discourse entities are more salient or accessible when described by a proper name than when described by a role (a definite description). One of the experiments assessed the effect of manner of mention (i.e. name or role) on the time taken a) to read a target sentence containing an anaphoric pronoun, and b) to answer a question referring to one of the discourse entities either by name or by role:

- (3.12) Mr. Bloggs/The manager was dictating a letter.  
Clare/The secretary was taking shorthand.  
It was getting to be late in the afternoon.  
He/She was beginning to feel hungry. (Target sentence)  
Question: Was Clare/the secretary carrying out filing?

Sanford *et al* found that reading times for targets introduced or referred to by proper names were significantly shorter than for those introduced or referred to by role descriptions.

According to Ariel, however, referring expressions do not determine the relative degree of accessibility of mental representations, rather the relative degree of accessibility of a representation (partially) determines the referring expression to be used. In other words, whilst various factors (such as those mentioned above) affect the degree of accessibility of mental representations relative to each other, these relative degrees of accessibility are encoded by different markers of accessibility. How then are Sanford *et al*'s results to be explained? The explanation lies, I suggest, largely in the fact that the test materials were not genuine discourses; in (3.12) above, neither Mr. Bloggs nor Clare were previously known to the subjects, and so the referring expressions used did not in fact refer to genuine discourse entities. The

---

<sup>29</sup> To avoid confusion, Ariel's term 'mental representation' will be adopted to refer to representations of discourse entities, as distinct from the term 'conceptual representation' which I will continue to use to refer to what is manipulated by procedural information.

effect might arise not as a direct result of the referring expression used, but as a result of the fact that named entities tend to be discourse topics, and discourse topics are highly salient, whereas described entities are less likely to refer to discourse topics and hence are less salient. By using a name to refer to a previously unknown entity, the author of (3.12) gives the effect of introducing a discourse topic (this is a common stylistic device); because of the nature of discourse topics, the experimental subject is expecting there to be continuity of topic, and so is primed for the pronoun in the target sentence to refer to the named referent rather than to the described referent.

It appears, then, that linguistic systems exploit the fact that mental representations of discourse entities differ in their relative accessibility, by ranking referring expressions according to some kind of hierarchy; language users can then exploit the linguistic system in situations where mental accessibility is yet to be determined, for example at the start of narratives. According to Ariel (1991:449) referring expressions enter into the following accessibility hierarchy (abbreviated):

#### LOW ACCESSIBILITY

- Full name
- Definite description
- Last name
- First name
- Distal demonstrative (+ Modifier)
- Proximal demonstrative (+ Modifier)
- Stressed pronouns + Gesture
- Stressed pronouns
- Unstressed pronouns
- Zeros

#### HIGH ACCESSIBILITY

Markers of low accessibility, such as names, are typically highly informative, relatively long phonologically and easily refer uniquely, whilst markers of high accessibility are typically uninformative, phonologically attenuated, and do not refer uniquely; zeros are an extreme example of this (Ariel 1994:32). Ariel (1988:82-83) argues that this is to be expected given the principle of relevance: low accessibility markers are costly to process both linguistically and in terms of extending the utterance interpretation context to include encyclopaedic knowledge; high accessibility markers, on the other hand, require a small amount of processing effort to decode, and signal that the addressee is to use the currently most accessible utterance interpretation context in the search for an intended referent, thus reducing the processing effort involved in contextual extension. Thus, the principle of relevance predicts that speakers should use the highest possible accessibility marker in a given referring expression in order to minimise the addressee's processing effort.

Which one the speaker chooses will depend on her assessment of the relative accessibility of the appropriate mental representation in the addressee's memory, hence some free variation in use and interpretation of accessibility markers is to be expected (Ariel 1994:38).

For example, according to the Accessibility hierarchy, zeros are markers of higher accessibility than pronouns. The principle that speakers should use the highest possible accessibility marker predicts that a zero should be used in preference to a pronoun when coindexed with the subject of the clause in which it occurs, since this has the highest possible accessibility level of any discourse referent. Hence Accessibility theory predicts that pronouns must be free in their governing categories:

(3.13)            John<sub>i</sub> wants Ø<sub>i</sub> /him\*<sub>i</sub> to win the prize.

Similarly, unstressed pronouns are markers of higher accessibility than stressed pronouns. The principle that speakers should use the highest possible accessibility marker predicts that if an unstressed pronoun is used in a context where there are two highly accessible possible discourse referents, it should be coindexed with the more accessible referent, but that if a stressed pronoun is used, it should be coindexed with the less accessible referent. This prediction is born out in the following examples (from Ariel 1988:79) in which 'Jane' is more accessible than 'Mary', being topical (recency of mention is not an overriding factor here since both 'Jane' and 'Mary' occur within the same clause):

- (3.14) a.        Jane<sub>i</sub> kissed Mary<sub>j</sub> and then she<sub>i</sub> kissed Harry.  
          b.        Jane<sub>i</sub> kissed Mary<sub>j</sub> and then SHE<sub>j</sub> kissed Harry.

[Ariel 1988 (17a) and (16a)]

In summary, as the accessibility hierarchy makes clear, pronouns are markers of relatively high accessibility (with further distinctions made in English through intonation and in languages like Kiswahili through the alternation between free and bound pronominal morphemes). A pronoun will therefore be used only when a speaker judges that the mental representation of the entity to which it refers is highly accessible to the addressee, relative to alternative mental representations. This claim is discussed in further detail below, with respect to the Kiswahili object marker.



### 3.1.3 The Kiswahili Object Marker as an exponent of procedural encoding

The procedural analysis of pronouns will be illustrated by reference to the Kiswahili 'object marker' (OM), the distribution and interpretation of which are conditioned by syntactic, semantic and discourse factors.

#### *Introduction*

Kiswahili is a noun class language; the noun classes manifest themselves morphologically in various ways: within tensed VPs, the noun class or person/number of any arguments may be indicated by the SP ('subject prefix') and OM ('object marker'). In tensed, affirmative, indicative verb complexes the following 'slots' are filled (obligatorily in bold, optionally in normal type):

SP | tense | relative pro. | OM | verb stem | suffix(es) | indicative

In addition to these morphologically bound forms, there are four types of morphologically free pronominals: emphatic personal pronouns (see table 1, below), and three types of emphatic demonstrative pronouns (roughly, 'that/those already referred to', 'this/these here' and 'that/those there') in concord with the noun classes:

Table 1: SP and OM concords; emphatic personal pronouns

Person	Noun Class	SP	OM	Emphatic
1st sing.		ni-	-ni-	mimi
1st pl.		tu-	-tu-	sisi
2nd sing.		u-	-ku-	wewe
2nd pl.		m-	-wa-	ninyi
3rd sing.	1	a-	-m(w)-	yeye
3rd pl.	2	wa-	-wa-	wao
	3	u-	-u-	
	4	i	-i-	
	5	li-	-li-	
	6	ya-	-ya-	
	7	ki-	-ki-	
	8	vi-	-vi-	
	9	i-	-i-	
	10	zi-	-zi-	
	11/14	u-	-u-	
(infinitive)	15	ku-	-ku-	
(locative)	16	pa-	-pa-	
(locative)	17	ku-	-ku-	
(locative)	18	mu-	-mu-	

As table 1 illustrates, there are two partially overlapping systems in evidence: one based on person and number with corresponding emphatic personal pronouns, and another based on the noun classes (with corresponding referential, proximal and distal

demonstratives, not represented); these systems overlap in the case of the 3rd person and class 1/2.<sup>30</sup>

The SP and OM take precedence over syntactic considerations (the relative order of morphologically free lexical items) in determining the proposition expressed. The following examples from Khamisi (1972:8) illustrate this (subscripted numbers on nouns indicate noun classes):

- (3.15) a.      nyundo    i + me + li + vunj + a      jiwe  
                  hammer<sub>9</sub> SP9 PERF OM5 break INDIC stones<sub>5</sub>  
                  *the hammer broke the stone*
- b.      nyundo    li + me + i + vunj + a      jiwe  
                  hammer<sub>9</sub> SP5 PERF OM9 break INDIC stones<sub>5</sub>  
                  *the stone broke the hammer*

### *The 'subject prefix'*

As table 1 illustrates, the SP and OM for each noun class except class 1 are phonologically identical; morpho-syntactically, however, they are distinct. Before discussing the OM in any detail I shall briefly describe the factors affecting the distribution of the SP. The SP, despite its name, need not agree with the semantic subject of the clause in which it occurs. In informal conversation, where there is sufficient discourse support for the intended interpretation, the SP may agree with the semantic object. This occurs in the case of quasi-passives (Russell 1985; see also Whiteley & Mganga 1969) which "can be considered as the result of topicalising, by subjectivising NP2 in NP1-V-NP2 structures, but without the formal trappings of passivisation i.e. there is no passive marker -w- in the verb, and no preposition *na* to mark the agent" (Russell 1985:478). (3.16) below is an example of a quasi-passive; the SP *u-* is in concord with the preverbal NP *wimbo*, which is a topicalised semantic object (the semantic subject, *watu mia*, occurs after the verb):

- (3.16)            wimbo u + ta + imba watu mia  
                  song<sub>11</sub> SP11 FUT sing people hundred  
                  *A hundred people will sing the song (it will be a special performance).*

SP agreement can thus be said to be with the topic (defined loosely as prominent information); it is no surprise that in practice the SP usually agrees with the subject, since subject has been defined as an aggregate function formed from topic and agent

---

<sup>30</sup> Classes 1 and 2, 3 and 4, 5 and 6, 7 and 8, and 9 and 10 correspond to singular and plural distinctions; the plural of those members of class 11/14 which are not abstract nouns are in class 10. There is one further OM, the reflexive prefix -*ji-* which is not in concord with any noun class.

(Givón 1976:152, Comrie 1981:101). Locatives can also function as topics, as in (3.17) below:

- (3.17) a. Nyumba + ni pale pa + li + fika mgeni  
 House LOC<sub>16</sub> that<sub>16</sub> SP16 PAST arrive guest<sub>1</sub>  
*At that house arrived a guest*
- b. Pa + li + fika mgeni nyumba + ni pale  
 SP16 PAST arrive guest<sub>1</sub> house LOC<sub>16</sub> that<sub>16</sub>  
*There arrived a guest at that house*

but Givón's claim (1976:174) that class 16/17/18 (locative) agreement is obligatory with preposed locatives is contradicted by the acceptability<sup>31</sup> of (3.18) where the SP agrees with the noun class of the agent:

- (3.18) Nyumba + ni pale a + li + fika mgeni  
 House LOC<sub>16</sub> that<sub>16</sub> SP1 PAST arrive guest<sub>1</sub>  
*At that house a guest arrived*

Whatever it agrees with, the SP is obligatory, occurring either with or without an overt coreferential NP. It can be unrealised when the MTA marker *-ka-* is used; the most common use of *-ka-* is in narrative to mark an event as subsequent to and often contingent upon a previous event. When discourse structure constrains the context to a sufficient degree, such that the subject of a *-ka-* marked clause is so highly accessible that it need not be overtly marked, the SP may be unrealised. In the following narrative, the omitted SP (*a-*) refers to the topic, 'Seif', who is the most recently mentioned entity in each instance:

- (3.19) Wa + li + kuta Seif Ø + ka + funga mlango wake bado na i + li + onekana  
 SP2 PAST find Seif - ASP close door<sub>3</sub> his still and SP9 PAST seem  
 ha + u + ta + funguliwa tena. Pili a + li + gonga mlango kwa nguvu lakini  
 NEG SP3 FUT be opened again Pili SP1 PAST knock door<sub>3</sub> in strength but  
 Seif Ø + ka + nyamaa makusudi kama yeye ni kiziwi. Wa + li + ita tena  
 Seif - ASP be silent intention as if he is deaf SP2 PAST call again  
 lakini Seif ndio kwanza kama Ø + ka + zibwa masikio kwa pamba.  
 but Seif it is first as if - ASP be closed ears with cotton  
*They found Seif's door still locked and it seemed it would remain like that. Pili knocked hard on the door but Seif pretended to be deaf. They called (him) again and again but Seif remained quiet as if someone had blocked his ears with cotton wool.*

(NS p.6, cited Moshi 1985:60-61)

<sup>31</sup> According to Bwana Amcir Haji of the Taasisi ya Kiswahili na Lugha za Kigeni, Zanzibar.

In all the examples cited in Moshi (1985) of VPs of the form  $\emptyset + ka + verb$  as well as in other instances not cited in Moshi (1985) the 'missing' SP is the third person singular/class 1 prefix *a-*. This, together with the fact that other phonological transformations occur in the same morphological environment (e.g. *ni + na + verb* is often realised as *na + verb*) suggests that, rather than being morphologically absent, the SP may simply be phonologically unrealised according to the following morpho-phonological elision rule, which allows the SP *a-* to be optionally realised as  $\emptyset$ :

$$a \text{ (morph}_{SP1}) \rightarrow \emptyset - ka \text{ (morph}_{MTA})$$

When this option is exercised depends primarily on discourse factors. To summarise, the SP is obligatory, but may be phonologically unrealised in certain contexts; it can occur either with or without an overt coreferential NP; and this NP is a topic, which need not be the semantic subject of a clause.

#### *A grammatical analysis of the OM*

In contrast to the SP, which is obligatory, the OM is usually absent, except where the object of the clause is animate (Maw 1976:399) - i.e. when the OM is one of the personal pronominal forms, including those in the third person, which correspond to noun class 1/2. An OM is obligatory, however, when the semantic object has been topicalised and occurs in a pre-verbal position:

(3.20) Kikombe, *ni + li + ki + vunja*  
 cup<sub>7</sub> I PASTOM<sub>7</sub> break

*The cup, I broke it*

(3.21) \* Kikombe, *ni + li + vunja*

When the OM does occur, either with or without an overt object NP, its occurrence is constrained by various syntactic and discourse factors. Various characterisations of the OM have been proposed in order to account for its distribution; below I shall briefly discuss three differing accounts of the morpho-syntactic status of the OM.

According to Moshi (1985) there are only two grammatical roles in Kiswahili - subject and object - and these are determined with reference to the SP and OM as follows. A subject is defined either as the NP in concord with the SP, or, in the absence of an overt NP, as the SP itself. A 'subject NP' can be either pre- or post-verbal, but the pre-verbal position dominates. Similarly, Moshi defines an object either as the NP in concord with the OM, or, in the absence of an overt NP, as the OM

itself. An 'object NP' can be both pre- and post-verbal, but the post-verbal position dominates.

Examples of objects according to Moshi's definition are the underlined expressions in the examples below (N.B. the suffix *-a* is the indicative suffix):

(3.22) Ni + li + m + l + ish + i + a mwanamke watoto  
I PAST OM1 eat CAUS PREP woman<sub>1</sub> children<sub>2</sub>

*I fed the children for the woman*

(3.23) ni + li + wa + l + ish + a watoto  
I PAST OM2 eat CAUS children<sub>2</sub>

*I fed the children*

(3.24) Ni + li + wa + l + ish + a  
I PAST OM2 eat CAUSATIVE

*I fed them*

(3.20) Kikombe, ni + li + ki + vunj + a  
cup<sub>7</sub> I PAST OM7 break

*The cup, I broke it*

(3.25) Ni + li + vunj + a kikombe  
I PAST break cup

*I broke a cup*

The NP *mwanamke* is in concord with the OM *-m-* (class 1) in (3.22) and is therefore the object according to Moshi's definition; the NP *watoto* (class 2) is not the object in (3.22) since it is not in concord with an OM (Kiswahili does not permit more than one OM in a single VP), but in (3.23) *watoto* is in concord with the OM *-wa-* (class 2) and is therefore the object. There is no overt NP in concord with the OM *-wa-* in (3.24), so the OM itself is the object. The object NP in (3.20), *kikombe*, is pre-verbal, but in (3.25) the absence of the OM *-ki-* means that *kikombe* is not the object.

I shall reject Moshi's analysis on the following grounds: First, whilst the OM is morphologically unitary in that its position in the VP and its phonological realisation remain constant, Moshi posits a distinction between the OM as object (3.24) and the OM as agreement marker [(3.22) (3.23) and (3.20)]. Secondly, Moshi's analysis allows the NP *kikombe* to be the object in (3.20) but not in (3.25), where it is unspecified for a grammatical relation; what the grammatical status of *kikombe* is in (3.25) is unclear.

In contrast, Bresnan & Mchombo (1987:15) claim that "the OM, which at first glance looks like an object agreement marker, is actually an incorporated object pronoun." In

conjunction with the uniqueness condition of Lexical Functional Grammar (LFG), which allows only one instance of a given grammatical role within a single clause, this analysis entails that in the absence of an OM an object NP will genuinely be an object, but that when an OM is present such a NP will be something other than an object. Bresnan & Mchombo (1987:15-6) suggest that a NP in concord with an OM is a floating topic<sup>32</sup> NP to which the OM is anaphorically linked.

They argue that the object NP in the absence of an OM is obligatorily post-verbal so as to maintain the order <Verb Object> (in fact the order is <Tense Object> since the OM, which is an incorporated object pronoun according to Bresnan & Mchombo, occurs after the MTA prefix but before the verb stem):

(3.25) c.        Ni + li + vunj + a kikombe    <Tense Object>  
                   I   PAST break    cup  
                   *I broke a cup*

(3.21)        \* Kikombe, ni + li+ vunja    \* <Object Tense>

Topic NPs (in concord with an OM), on the other hand, are free of such word order restrictions, being outside the clause; topic NPs and subject NPs are optional and unordered with respect to each other and to the VP, hence word order is free within a sentence containing an OM.<sup>33</sup>

A morpho-syntactically unitary characterisation of the OM is also proposed by Seidl & Dimitriadis (1996), but whereas Bresnan & Mchombo (1987) characterise the OM as an incorporated object pronoun, Seidl & Dimitriadis characterise it as an inflectional affix (that is, an agreement marker) capable of licensing null objects. They argue from the principle of lexical integrity that if the OM is an incorporated pronoun, then the SP and MTA markers must also be clitics; this, they claim, is implausible since the SP and MTA marking are obligatory and therefore part of the verb. Seidl & Dimitriadis (1996) argue that if the OM is an inflectional affix, then co-occurring object NPs occur in verb complement position, whilst object marked verbs without overt object NPs are instances of pro-drop, that is, they have a null object as the argument of the verb.

---

<sup>32</sup> The term 'topic' has already been used in relation to SP agreement; however, as used in Bresnan & Mchombo (1987) 'topic' refers to an NP in concord with the OM. Henceforth, all references to 'topic' assume this latter usage of the term.

<sup>33</sup> A similar analysis is given in Allan (1983) where it is claimed that the OM is always a pronoun and in particular that, when coreferential with an overt NP, the OM is a cataphoric resumptive pronoun.

Seidl & Dimitriadis claim that a null object can only occur in the presence of an OM, and hence that an OM is required to license null objects. For a verb to be considered genuinely transitive in Seidl & Dimitriadis (1996), it had to meet any of the following criteria: 1) there was an overt object either following or preceding the verb, 2) there was an OM in the VP, 3) in English, the verb could not appear without an object (which would show that an object was not semantically required). However, examples of transitive verbs used without either an overt object NP or an OM are common in spoken Kiswahili:

- (3.26) A Umeleta chakula? *Have you brought (the) food?*  
 B Ni + me + leta Bwana. *I have brought it, sir.*  
 1sing PERF bring (Ashton 1947:45)
- (3.27) A Je, una watoto? *Do you have (any) children?*  
 B Bado. *Not yet.*  
 A U + ta + pata *You will do. (Lit. You will get.)*  
 2sing FUT get
- (3.28) A Chai na sukari? *Tea with sugar?*  
 B Si + tumi + i *I don't take sugar. (Lit: I don't use.)*  
 1sing use NEG

The examples above are clearly transitive by criterion 3, yet they do not contain either OMs or overt object NPs, hence it is not the case (as Seidl & Dimitriadis claim) that null objects must be licensed by OMs. Because the discourse referents in the above examples are very highly accessible (being explicitly mentioned in the preceding clauses) markers of very high accessibility (i.e. zeros) may be used.

In addition, both Seidl & Dimitriadis (1996) and Bresnan & Mchombo (1987) fail to account for the fact that it is not always necessary to have a coreferential OM in order to topicalise a semantic object. When there is an animate indirect object, and an inanimate direct object which is also a discourse topic, the direct semantic object can be topicalised even when the OM is in concord with the animate indirect semantic object, as in (3.29B) (A establishes *kitabu* as the discourse topic):

- (3.29) A Kitabu nilichokuwa nasoma jana kimeibiwa, sikioni.  
*The book which I was reading yesterday has been stolen, I can't see it.*  
 B Kitabu<sub>7</sub>, Juma alim<sub>1</sub>pa Mariamu<sub>1</sub>, hakikuibiwa.  
*The book, Juma gave it to Mariamu, it hasn't been stolen.*

Furthermore, some informants rejected (3.30a) in which the OM is in concord with the topicalised NP (see below for discussion of animacy and informant judgements):

- (3.30) a. \* Kitabu<sub>7</sub>, Juma aliki<sub>7</sub>pa Mariamu<sub>1</sub>, hakikuibiwa.  
 b. \* Kitabu<sub>7</sub>, Juma alipa Mariamu<sub>1</sub>, hakikuibiwa. (no OM)

In the above examples, it appears that an OM must be present in order for the NP *kitabu* to be topicalised, but that this OM need not be in concord with the topicalised NP (indeed cannot be when the topicalised NP is an inanimate semantic direct object). One possible explanation of this constraint is that Kiswahili only recognises one grammatical object (as Moshi 1985 claims), and that in double object constructions this is the indirect object (see below for further evidence for this claim). The semantic direct object is therefore not a grammatical object at all, and so is free of the grammatical <Subject Tense Object> word order restriction and can, if the discourse supports it, be topicalised in the absence of a coreferential OM.

Leaving aside this consideration, it is difficult to distinguish empirically between Seidl & Dimitriadis' analysis of the OM as an inflectional affix and Bresnan & Mchombo's analysis of the OM as an incorporated object pronoun. The differences stem from the fact that Seidl & Dimitriadis (1996) assumes a Principles and Parameters framework while Bresnan & Mchombo (1987) assume LFG. Which analysis is to be preferred depends largely on which framework is adopted. However, I shall adopt Bresnan & Mchombo's terminology, and call the OM an incorporated pronoun, since, as Seidl & Dimitriadis found (in line with Hyman & Duranti 1982 and Allan 1983), the distribution of the OM resembles that of a pronoun rather than an inflectional affix. For example, the OM need not agree with the noun class of the NP with which it is anaphorically linked:

- (3.31) a. A + li + ki + ona kitoto  
 SPI PAST OM<sub>7</sub> see baby<sub>7</sub>  
 b. A + li + mw + ona kitoto  
 SPI PAST OM<sub>1</sub> see baby<sub>7</sub>  
*She saw a/the baby*

In (3.31a) the agreement is based on noun class, but in (3.31b) it is semantic - class 1 being the class of singular animates. This phenomenon is discussed further in Reynolds & Eastman (1989).

#### *The OM in discourse*

Aside from syntactic constraints, the distribution of the OM is constrained by semantic and discourse factors. Like unstressed personal pronouns in English, the SP and OM are both markers of high accessibility. This can be demonstrated in



comparison with emphatic personal pronouns<sup>34</sup> (parallel to the comparison of unstressed and stressed personal pronouns in English):

- (3.32) a. Moja<sub>i</sub> a<sub>i</sub>+li+m<sub>j</sub>+kumbatia Pili<sub>j</sub>, halafu a<sub>i</sub>+li+m<sub>k</sub>+kumbatia Tatu<sub>k</sub>.  
*Moja<sub>i</sub> hugged Pili<sub>j</sub>, then he hugged Tatu<sub>k</sub>.*
- b. Moja<sub>i</sub> a<sub>i</sub>+li+m<sub>j</sub>+kumbatia Pili<sub>j</sub> halafu yeye<sub>j</sub> a<sub>j</sub>+li+m<sub>k</sub>+kumbatia Tatu<sub>k</sub>.  
*Moja<sub>i</sub> hugged Pili<sub>j</sub> then HE<sub>j</sub> hugged Tatu<sub>k</sub>*
- (3.33) a. Moja<sub>i</sub> a<sub>i</sub>+li+m<sub>j</sub>+kumbatia Pili<sub>j</sub>, halafu Tatu<sub>k</sub> a<sub>k</sub>+li+m<sub>j</sub>+kumbatia<sup>35</sup>  
*Moja<sub>i</sub> hugged Pili<sub>j</sub>, then Tatu<sub>k</sub> hugged him<sub>j</sub>*
- b. Moja<sub>i</sub> a<sub>i</sub>+li+m<sub>j</sub>+kumbatia Pili<sub>j</sub> halafu Tatu<sub>k</sub> a<sub>k</sub>+li+m<sub>i</sub>+kumbatia yeye<sub>i</sub>;  
*Moja<sub>i</sub> hugged Pili<sub>j</sub>, then Tatu<sub>k</sub> hugged HIM<sub>i</sub>*

A discourse referent can become highly accessible as the result of various cognitive and discourse processes (as mentioned in §3.1.2). For example, definites tend to be more accessible than indefinites, and animates tend to be more accessible than inanimates. It has been widely claimed that the OM marks definite objects, and is obligatory with animates, but both these claims are incorrect. I shall address each in turn before discussing those factors which do affect the distribution of the OM and increase the relative accessibility of discourse referents.

The claim that the OM marks definites is common (cf. Wilson 1970:186, Driever 1976:23, Givón 1976, Hinnebusch 1979:218, Rugemalira 1991:207), however Wald (1979) cites examples where an OM is present but the coreferential NP is indefinite (and in the case of (3.35) below nonreferential also):

- (3.34) akam<sub>1</sub>kuta mzee<sub>1</sub> mwingine ndugu wa yule  
*then she met another old lady, sister of the first one* (Wald 1979:512)
- (3.35) nao mahala wanapoweza kutoa lile dukuduku ni kum<sub>1</sub>piga mtu<sub>1</sub>  
*and the only way they can get rid of their frustration is to hit somebody*  
 (ibid. 513)

Conversely, Allan (1983:325) cites examples of definite NPs (underlined) with no coreferential OM:

- (3.36) Huwezi kununua motokaa hiji bila fedha nyingi.  
*You can't buy this car without a lot of money.*

<sup>34</sup> Emphatic personal pronouns can occur alone, but when occurring in a sentence a corresponding SP or OM must be present, entailing that, on the present analysis, emphatic personal pronouns are always topic NPs.

<sup>35</sup> This is the interpretation of Kassim Abdullah (Queen's University Ontario), whilst Rchema Rajabu (University of York, England) found this example ambiguous.

- (3.37) Asma alishonesha nguo zake Nairobi.  
Asma had her dresses sewn in Nairobi.
- (3.38) Tulipokwenda Dar tulitembelea chuo kikuu.  
When we were in Dar we visited *the*/\*a university. (there is only one)

Animacy has also been proposed as a determining factor in the occurrence of the OM. Wald (1979:508) and Rugemalira (1991:204) argue that 1) only an animate NP may be object marked when there is more than one semantic object to the verb, and 2) that if both NPs are animate, the prepositional or indirect object must be marked (Kiswahili does not allow multiple object marking):

- (3.22) a. ni + li + m + l + ish + i + a    mwanamke    watoto  
SP1 PAST OP1 eat CAUS PREP    woman<sub>1</sub>        children<sub>2</sub>  
*I fed the children for the woman.*
- b. \*Ni + li + wa + l + ish + i + a    mwanamke    watoto  
SP1 PAST OP2 eat CAUS PREP    woman<sub>1</sub>        children<sub>2</sub>

The first constraint may well reduce to the second since in double object constructions with animate and inanimate NPs, the animate NP is usually the indirect object since its referent is likely to be a recipient or beneficiary. The second constraint also provides support for the claim made above that Kiswahili only allows one grammatical object per clause, and that this will be the semantic indirect object in double object constructions.

Allan (1983:332) goes further in claiming that the OM is obligatory with animate coreferential NPs of class 1/2<sup>36</sup> even when the NP is indefinite and nonreferential (this is widely taught as a grammatical 'rule'):

- (3.39) a. Simw<sub>1</sub>oni mtu<sub>1</sub> yeyote.        *I don't see anyone.*  
b. ? Sioni mtu yeyote.
- (3.40) a. Sik<sub>1</sub>oni kitu<sub>1</sub> chochote.        *I don't see anything.*  
b. Sioni kitu chochote.

The problem with Allan's stronger claim is that informant judgements are inconsistent and often at odds with actual usage where animacy is concerned. For

---

<sup>36</sup> The exception being when "the focus of information falls on the predicate, such that the combination of Verb and ONP [object NP] has a semantic unity reminiscent of phrasal verbs." (Allan 1983:332) E.g.:

- (6) a. Nilimwona mganga.        *I saw the doctor.*  
b. Niliona mganga.        *I sought medical advice.*

example, while one informant rejected (3.41b) with the OM coreferential with an inanimate NP, Khamisi (1972:15) has both (3.42a) and (3.42b), where the latter has an OM in concord with *maziwa* ('milk') rather than with *Saidi* (a person and therefore class 1); (3.42b) is the marked form, however, and is only acceptable in a context where there is contrastive emphasis on *maziwa*.

- (3.41) a. Kitabu<sub>7</sub>, Juma alim<sub>1</sub>pa Mariamu<sub>1</sub>.  
*The book, Juma gave it to Mariamu.*  
 b. \* Kitabu<sub>7</sub>, Juma alik<sub>1</sub>pa Mariamu<sub>1</sub>.
- (3.42) a. Musa amem<sub>1</sub>letea Saidi<sub>1</sub> maziwa<sub>5</sub>.  
*Musa brought Saidi some milk.*  
 b. Musa ameya<sub>5</sub>letea Saidi<sub>1</sub> maziwa<sub>5</sub>.  
*Musa brought the milk (not the water etc.) for saidi.*

Corpus based analyses confirm that, although there is a tendency for animate objects to be object marked, the occurrence of an OM is not obligatory: for example, Seidl & Dimitriadis (1996) found that whilst 104 animate NPs were object marked in their data of modern narrative texts, 40 were not. Rather, the conditioning factor on the occurrence of an OM is that the referent of the OM must be highly salient, where salience is defined in terms of frequency of mention (Moshi 1985) or recency of mention (Seidl & Dimitriadis 1996). Saliency, being a relative term, is further defined in comparison to entities overtly marked by object NPs alone. This finding corresponds with a characterisation of the OM as a marker of relatively high accessibility compared with object NPs; mental representations of entities which have been mentioned frequently or recently will be more accessible to an addressee than those which have not, therefore a high degree of salience as measured by Moshi (1985) and Seidl & Dimitriadis (1996) correlates with high relative accessibility as given in Ariel's (1991) accessibility hierarchy.

That animates tend to be more salient than inanimates is uncontroversial: in one narrative (from MWK) Moshi (1985:73) found two humans both mentioned 51 times, whilst the next most salient referent (*chembe* 'match') was mentioned only nine times. The effects of definiteness and animacy on the distribution of the OM can be subsumed under the relation of the OM to highly accessible, or salient, discourse entities, as Moshi (1985:79) explains:

"Because discourses are overwhelmingly about animates (human mostly) the tendency for more animate than inanimate nominals to appear in discourse [i.e. to be salient] and for these animates to be definite is not a surprise to us. It is a reflection of discourse tendencies rather than a general rule of grammar to which speakers have to abide when constructing their discourses."

Given this, it is more plausible to claim that animates in discourse tend to be topic NPs, that is, representations of highly accessible entities (a discourse notion) rather than, as Moshi (1985) claims, that animate as opposed to inanimate NPs tend to be grammatical objects (a discourse independent notion).

This said, the high frequency of animate (class 1/2) OMs can also be related to the claim that Kiswahili allows only one grammatical object per clause, and that this must be the indirect semantic object in double object constructions. Since the OM is itself always a grammatical object whether it occurs with or without an overt coreferential NP, it is to be expected that the OM will be coreferential with the indirect semantic object when there is a choice. Since animates tend to be indirect objects and also tend to be more salient than inanimates, grammatical and discourse constraints work together to favour the occurrence of animate (class 1/2) OMs over other, inanimate OMs.

Occasionally, however, grammatical and discourse constraints are at odds, for example when the direct semantic object is highly salient. In such cases, informant acceptability judgements vary: apparently some speakers allow discourse constraints to override grammatical constraints (and therefore find (3.42b) acceptable, given an adequate supporting context) whilst others maintain the grammatical constraint on indirect objects only being object marked (as in (3.41b)).

### *Summary*

The distribution and interpretation of the OM can be accounted for if the OM is viewed as an exponent of procedural encoding. The procedural information encoded by the OM relates both to noun class and to accessibility. The OM marks highly accessible discourse entities, as the contrast between the OM and emphatic personal pronouns in example (3.33) demonstrates. Because the OM marks highly accessible discourse entities, these must be part of the GROUND (that is old or background information) rather than the FOCUS (new or foreground information) in the sense of Vallduvf's information packaging (see §2.2.2). The following examples (repeated from §2.2.2) illustrate this.

- |                       |    |   |
|-----------------------|----|---|
| (2.33)                | A: | <i>What did you do?</i>   |
|                       | B: | ni + li + vunja kikombe<br>SP PAST break cup <sub>7</sub><br><i>I broke a cup</i> |
| formational structure |    | ALL-FOCUS   |

All of the information conveyed by B's utterance in (2.33) is new, so the formational structure is ALL-FOCUS; the cup is not an accessible discourse entity, since its first mention is by B, and so there is (and can be) no OM. In the following examples, the cup is highly accessible, having been previously mentioned; a corresponding class 7 OM is therefore possible, either with or without an overt NP.

- (2.34)           A:    *What did you do to the cup?*  
                   B:    ni + li + ki + vunja  
                           SP PAST OM7 break  
                           *I broke it*  
 formational structure    ALL-FOCUS

The OM in (2.34) is not a LINK because it is assumed that the addressee is already at the appropriate address, that is, the cup is highly accessible; the OM therefore plays a similar informational role to pronouns in English, which is to be expected given the above characterisation of the OM as an incorporated object pronoun.

- (2.35)           A:    *What happened to the cup?*  
                   B:    (kikombe) ni + li + ki + vunja  
                           cup7            SP PAST OM7 break  
                           (*the cup*)    *I broke it*  
 formational structure    (LINK)            FOCUS

- (2.36)           A:    *What happened to the cup?*  
                   B:    ni + li + ki + vunja (kikombe)  
                           SP PAST OM7 break cup7  
                           *I broke it        (the cup)*  
 formational structure    FOCUS                (TAIL)

In both (2.35) and (2.36) the object marked VP *nilikivunja* (I broke it) is the FOCUS since it constitutes new information not presupposed by A's previous utterances. Although it is usually the FOCUS which, because of its informational importance, is emphasised, elements of the GROUND may also receive emphasis. In contrast to the OM, the use of the overt NP *kikombe* (the cup) is optional (as the brackets indicate), but its occurrence in conjunction with the corresponding OM has the effect of redirecting the addressee towards the currently most accessible address and hence emphasising it. Additionally, the OM allows the overt NP to achieve syntactic prominence, since only if an OM is present can the NP function as a syntactic topic which, being free of restrictions on word order, is able to be topicalised, as in (2.35) where the NP *kikombe* occurs in the highly salient sentence initial position.

### 3.1.4 Conclusion

Referring expressions other than pronouns encode conceptual information; for example, a proper name gives access to a concept the content of which includes encyclopaedic knowledge relating to the individual denoted by that name, and descriptions consisting of a determiner (which is procedural - see §2.1.2) and a NP encode conceptual information in the NP, such as the concept {DOG} to which the NP *dog* gives access in the phrase *the dog*. Pronouns, unlike other referring expressions, encode only procedural information; that is, they encode processing instructions by which an addressee can identify an intended referent, but they do not give rise to any conceptual information.

According to the principle of relevance, addressees aim to interpret a given utterance as efficiently as possible; for this reason, addressees will, in the first instance, process new information within the currently most accessible utterance interpretation context, and will only extend or revise this if adequate contextual effects can not be computed, or if there is some indication that the current context will prove insufficient (for example through the use of the discourse connective *well* - see §2.2.1). Pronouns reflect and exploit this procedure by constraining the addressee's search for intended referents to the set of highly accessible discourse entities. Characterisation of the Kiswahili OM in this way accounted neatly for its distribution and discourse functions. Individual pronouns in different languages also narrow the search by encoding information relating to one or more of number, person, gender, case, noun class, etc. The use of a pronoun over a marker of lower accessibility also reduces processing effort at phase one, where the frequency of mention and phonologically and semantically reduced nature of pronouns renders them easy to decode.

In the following section, I shall consider other types of linguistically encoded information which could be considered to encode procedural information at phase two of utterance interpretation. Given the parallels drawn in the literature between pronominal and temporal anaphora (Partee 1973, 1984; Carston 1988b) and between pronominal anaphora and modality (Roberts 1989), I shall extend the discussion of procedural encoding at phase two to incorporate tense, aspect and modality markers which, like pronouns, are indexical expressions which help determine propositions. Whereas pronouns constrain the manipulation of sub-propositional conceptual representations of entities projected from the lexical entries of predicates, tense, aspect and modality markers constrain the manipulation of sub-propositional conceptual representations of situations or events. The way in which they do this and the aspects of the utterance interpretation context whose construction is thereby constrained are discussed below.

## **3.2 Situations and Events**

### **3.2.0 Introduction**

All of the examples of procedural encoding which we have so far considered (at phases two and three) have had the effect of constraining the hypothesis space to be searched by the addressee during the inferential phases of utterance interpretation. Procedural information functions, I have claimed, by relating some conceptual representation to a specific aspect of the utterance interpretation context. In the case of a nominal referring expression, the hypothesis space is constrained in regard to the supposed relative degree of accessibility (in the mind of an addressee) of the mental representation of the intended referent.

However, it is not only mental representations of discourse entities, with varying degrees of relative accessibility, that the utterance interpretation context contains, but also representations of the situations or events in which these entities participate. An addressee entertains assumptions concerning, for example, the existential and temporal status of these situations and events, that is, whether these are to be represented as actual (i.e. verified or verifiable) or 'irrealis' (i.e. unverified and hence potential, hypothetical, conditional, etc.), and as occurring - in reality or potentially - at some time. This suggests that modality-tense-aspect<sup>37</sup> (MTA) markers might constitute a class or classes of procedurally encoded expressions. If this were the case, MTA markers would constrain representations of situations and events with respect to modal, temporal and aspectual features of contexts, in a similar way to that in which pronouns and other types of referring expression constrain the identification of discourse entities according to the supposed accessibility of their mental representations in the mind of an addressee.

Take, for instance, the temporal reference of (3.42):

(3.42)           Jane kissed Mary.

Both 'Jane' and 'Mary' encode concepts indicating that an instance of a particular individual is at stake, and 'kiss' also encodes a concept, expressing an event of a particular kind involving at least two individuals and occurring at a place and a time.

---

<sup>37</sup> In this I follow the 'natural' order identified by Bybee (1985), although Bickerton (1981, 1984) and Muysken (1981) claim that the usual order in (plantation) creoles is TMA, realised as [+anterior], [+irrealis], [+non-punctual] (a view challenged by Kihm's 1995 study of Tayo, a New Caledonian plantation creole). I shall also use the term 'modality' in preference to the more common 'mood' given that 'mood' has already been used with reference to sentence type (declarative, interrogative etc.) and also serves to distinguish verbal categories such as indicative and subjunctive.

In contrast, the past tense suffix, *-ed*, does not encode information about itself so much as about the interpretation of the event expressed by (3.42), namely that the event [Jane kiss Mary] is to be interpreted as occurring at some time in the past. This, I shall argue in §3.3, is procedural information. In the same way that in (3.13a) and (3.13b):

- (3.13) a. Jane<sub>i</sub> kissed Mary<sub>j</sub> and then she<sub>i</sub> kissed Harry.  
b. Jane<sub>i</sub> kissed Mary<sub>j</sub> and then SHE<sub>j</sub> kissed Harry.

reference assignment to 'she' and 'SHE' helps to determine the proposition expressed ((3.13a) is truth-conditionally distinct from (3.13b)), so the past tense suffix, *-ed*, helps determine the truth-conditions of (3.42) and therefore functions at phase two of utterance interpretation.

In order to provide a viable analysis of MTA markers in procedural terms, two things must be specified: the nature of the conceptual representations of situations and events which MTA markers manipulate, and those features of the utterance interpretation context in relation to which these representations are manipulated. I shall address each in turn.

### 3.2.1 Situation representations and operators

In this subsection, I propose a characterisation of sub-propositional conceptual representations of situations and events as 'situation representations', based on Klinge (1993). Klinge (1993:320-321) proposes a model in which a logical form consists of a 'propositional content' (PC) plus various 'operators';<sup>38</sup> these encode conceptual and procedural information respectively. Operators have the PC within their scope, as follows:

(3.43) OPERATORS [PROPOSITIONAL CONTENT] [Klinge's (17)]

Thus each of the sentences (from Klinge 1993:320):

(3.44) John buys Mary's ticket.

(3.45) John bought Mary's ticket.

---

<sup>38</sup> The use of the term 'operator' has been extended (on a par with the extension of this term in e.g. Cobb (1995) to discussions of scope within autosegmental phonology) to designate procedural information with scope over conceptual representations in the language of thought; its use here is not restricted to the first auxiliary in a verb phrase.



- (3.46) Did John buy Mary's ticket?  
 (3.47) John will buy Mary's ticket.

can be represented as:

- (3.48) OPERATORS [JOHN BUY MARY'S TICKET] [Klinge's (18)]

The PC, [JOHN BUY MARY'S TICKET] still needs to be inferentially enriched, for example to distinguish between possible interpretations of 'Mary's' such as 'for Mary', 'from Mary', etc. The inferentially enriched PC Klinge calls a 'situation representation' (SR). A SR is still not fully propositional, that is, a SR alone cannot be truth-conditionally determined since it is a representation of a situation or event devoid of modal, temporal, or aspectual information.<sup>39</sup>

However, the status of SRs is unclear in Klinge (1993) which presents two incompatible views of SRs. On the one hand Klinge (1993:320) suggests that SRs are propositional since they combine directly with elements of the context ({C}) to achieve contextual effects, that is, they provide the input {P} to phase three of utterance interpretation:

"... I will call the propositional form a sentence gives rise to a SITUATION REPRESENTATION.... It is the SITUATION REPRESENTATION, not the linguistic semantic input, which an addressee can combine with existing knowledge and assumptions about the state of affairs referred to to form new assumptions."  
 (Klinge 1993:320)

On the other hand, Klinge (1993:321) claims that SRs are derived solely from PCs (i.e. minus operators) and must therefore be sub-propositional:

"It is the PROPOSITIONAL CONTENT that undergoes enrichment, disambiguation and reference assignment to turn it into a conceptual representation of a SITUATION."  
 (Klinge 1993:321)

These characterisations of SRs are only compatible if operators do not contribute to propositions and thus affect truth conditions, in which case all procedural information would function at what I have termed phase three of utterance interpretation - as early work on the conceptual/procedural distinction (e.g. Blakemore 1987) assumed. This

---

<sup>39</sup> A similar assumption is made in DRT also. In DRT, truth-conditional interpretation only occurs once a DRS has been constructed, which involves anaphora resolution and specification of the temporal relations among the situations and events represented. Thus neither pronouns nor tense and aspect markers are directly truth-conditionally interpretable in DRT, rather, a discourse is truth-conditionally interpretable only via a series of DRSs (cf. e.g. Reyle & Gabbay 1994).

situation results from Klinge's conflation of the two inferential phases of utterance interpretation, (much as in Wilson & Sperber (1993a:1), see footnote 18); Klinge includes not only syntactic mood as an operator (as the non-declarative word order in example (3.46) above illustrates) but also tense and modality markers. Whilst syntactic mood functions at phase three of utterance interpretation and takes whole propositions within its scope (as, for example, does the Kiswahili interrogative *je* and the discourse connectives *so* and *well*), the use of a particular tense to locate an event in time contributes to the identification of a proposition, and therefore functions at phase two. The distinction between these two types of procedural information is lost by classifying them both as examples of a single undifferentiated semantic type: 'operator'.

In contrast to the wide range of elements which Klinge counts as operators, PCs consist of a highly restricted set of elements. The sense in which Klinge uses the term 'propositional content' (which is also the sense which this study will adopt) contrasts with the use of the same term in Wilson & Sperber (1988:133), although in both accounts propositional content is independent of context:

Figure 8: uses of the term *propositional content*

Klinge (1993) includes:	<i>operators</i> tense, modals, syntactic mood	<i>propositional content</i> 'content words'
Wilson & Sperber (1988) includes:	<i>propositional attitudes</i> syntactic mood	<i>propositional content</i> tense, modals, 'content words'

To avoid any terminological confusion, the use of the terms 'operator' and 'situation representation' in this thesis will be specified as follows. 'Operator' will be used exclusively to refer to linguistic elements (specifically modality, tense and aspect markers) functioning procedurally at phase two of utterance interpretation (that is, in a more restricted way than Klinge 1993). SRs are more complicated to characterise; this will be done first negatively in terms of what SRs are not, and then positively, stating what they are.

Firstly, a SR is distinct from the propositional form of an utterance (contra Klinge 1993:320), which consists of the inferentially enriched PC interpreted in accordance with the procedural information encoded by all of the associated operators. This is illustrated in the following example: the utterance (3.49a) is represented in the language of thought as (3.49b) which is a representation of the type OPERATOR [PC]; the PC is then inferentially enriched through reference assignment of the pronouns and

disambiguation of the genitive to give (3.49c) which is a representation of the type OPERATOR [SR]; the SR, which is not a propositional form, is then manipulated in line with the procedural information encoded by *will*<sup>40</sup> to give rise to the fully propositional (3.49d):

- (3.49) a. He will buy her ticket.  
b. WILL [HE BUY HER TICKET]  
c. WILL [JOHN BUY A TICKET FOR MARY]  
d. John will (at some future time) buy a ticket for Mary.

At the most basic level of definition, a SR is a conceptual representation of a situation or event which is manipulated by an operator relative to some feature of the utterance interpretation context. However, a SR need not be derived solely from the inferentially enriched PC (contra Klinge 1993:321), since operators differ with respect to their relative scope. In English, as in many other languages (Bybee 1985), modal markers have wider surface syntactic scope than temporal markers which in turn have wider surface syntactic scope than aspectual markers, and this is reflected in the language of thought by the relative scope of the corresponding operators:

- (3.50) a. Mary will have been working.  
b. WILL [ANTERIOR [PROGRESSIVE [SR MARY WORK]]]  
c. WILL [ANTERIOR [SR MARY BE WORKING]]  
d. WILL [SR MARY HAVE BEEN WORKING]

In (3.50b) the progressive aspect operator has scope over the SR [MARY WORK], derived from the PC [MARY WORK] inferentially enriched to identify the intended 'Mary'. The temporal (anterior) operator in turn has scope over the resulting SR (3.50c), and finally (3.50d), the modal operator has scope over the SR derived from the inferentially enriched PC, the aspectual operator and the temporal (anterior) operator.

This entails that a SR is only defined relative to the operator which has immediate scope over it. Therefore, whilst we can talk of the (unique) PC encoded by an utterance we cannot talk of the (unique) SR (unless only one operator is encoded in the utterance). The PC in any utterance consists purely of linguistically encoded conceptual information, whereas a SR within the scope of a given operator, O, consists of the conceptual information encoded by the PC inferentially enriched in line

---

<sup>40</sup> Arguments for viewing *will*, as an operator are provided in §6.1; for the moment this is simply assumed.

with the procedural information encoded by any other operators with narrower scope than O.

When a particular operator manipulates a SR, it does so in relation to some specific feature of the utterance interpretation context. This, the second element of a procedural account of MTA markers in need of specification, will be discussed below.

### 3.2.2 Domains

An utterance interpretation context is, as we have seen, a complex construction, consisting of various assumptions entertained with varying degrees of certitude. A SR may be interpreted relative to any of these assumptions at phase two of utterance interpretation, as may propositional forms at phase three. There are, however, features of the utterance interpretation context relative to which only the sub-propositional SRs may be interpreted, since they contribute directly to the recovery of propositional form. These are the features of context which specify existential, temporal and aspectual information, and it is in relation to these, which I shall term *domains*, that most MTA markers constrain the interpretation of SRs. The distinction between expressions which relate SRs to cognitive domains and those which relate SRs to other aspects of the utterance interpretation context will be important in distinguishing *will* and *shall* from the other modals. It will be shown in chapter 4 that *can*, *may*, *must* and *should* relate SRs to assumptions in the utterance interpretation context and in chapter 6 that *will* and *shall* relate SRs to a cognitive domain (*will* to the irrealis domain discussed below, and *shall* to the irrealis and temporal domains).

The term *domain* is derived from Langacker's use of this term within the framework of cognitive grammar (Langacker 1983:155-162).<sup>41</sup> According to Langacker (1983:155) "every predicate is characterized relative to one or more *domains*, collectively called its *matrix*." In my use of the term, it is contexts rather than predicates which are characterised relative to domains. In relation to the representation of situations and events, domains specify the existential, temporal, and aspectual parameters within which a SR can be interpreted; that is, domains specify whether an event is to be interpreted as verified or unverified, the time of (actual or potential) occurrence of an event, and the nature of that occurrence - whether it occurs once, continuously or iteratively, is completed or incomplete, etc. Only once values have been assigned to all of these parameters, i.e. when a SR has been related to all of the relevant domains (the matrix), can a propositional form be said to have been

---

<sup>41</sup> Klinge (1993:325-329), in line with Durst-Andersen's Mental Grammar (1986, 1992), suggests that operators relate SRs to various 'indices', which fulfil a similar role to that of domains in my account.

recovered. Domains can thus be viewed as providing the cognitive framework within which propositions are constructed.

The relating of SRs to domains is an inferential process occurring at phase two of utterance interpretation. In the absence of any explicit indication to the contrary, an addressee will interpret a SR relative to the currently most accessible domain. In the case of the existential status of an event the default assumption will usually be that the SR represents a verified rather than an unverified event (speakers are more likely to talk about events which they know have occurred or are occurring than about potential or hypothetical events). This default assumption can, however, be overridden; procedural information concerning the intended interpretation of a SR with respect to a given domain may be encoded by a speaker in the form of one or more operators. Hence, operators relating SRs to the realis (actual, verified) domain are rare, since this is taken as a default parameter; in contrast, operators relating SRs to the irrealis (potential, hypothetical) domain, such as markers of modality, are attested in the majority of languages (Bybee *et al* 1994).<sup>42</sup> This process parallels that which Ariel described with respect to nominal referring expressions and their relation to context:

"Note that by itself, Relevance predicts that a context search be conducted according to degree of context Accessibility, automatically giving priority to the more accessible context compatible with Optimal Relevance. Accessibility theory, however, offers a marking system which eliminates this default assumption in favour of a clearer instruction of the speaker to the addressee regarding 'where' to retrieve the background information required."

(Ariel 1990:171)

Domains can be characterised as either 'basic' or 'abstract'. Basic domains cannot be fully reduced to any other domain. For example neither the temporal domain nor the irrealis domain can be defined in terms of any other domain, and are therefore basic. A concept characterised relative to one or more basic domains can itself function as the domain for a number of other concepts, which can in turn function as domains for further concepts, and so on indefinitely. These non-basic domains are termed 'abstract domains'. For example, both the temporal domain and the irrealis domain are involved in the representation of future time reference, since a situation or event in the future is also logically irrealis, that is, it is non-actual and unverified. The future can therefore be viewed as an abstract domain defined relative to the temporal and the

---

<sup>42</sup> In positing a basic binary distinction between realis and irrealis I follow e.g. Chung & Timberlake (1985) who also make a distinction between 'realis' and 'irrealis', the latter including what Leech (1971:112) calls 'theoretical' and 'hypothetical' and what Matthews (1989:196) calls 'potentialis' (potential) and 'irrealis' (unreal). The alternative position (adopted by Leech 1971 and Matthews 1989) is to posit a basic three way distinction between real (factual), potential and hypothetical.

irrealis domains. The relation between the temporal and irrealis domains with respect to the future is illustrated in a simplistic form in figure 9, which shows the impossibility of an event or situation being both future and realis:

Figure 9:

past / realis	present / realis	
past / irrealis	present / irrealis	future / irrealis

This is reflected in many languages, such as Takelma (Sapir 1912) which combines a realis form of the verb stem with the non-future form of a suffix indicating person and number, and an irrealis form of the verb stem with the future form of a suffix indicating person and number:

(3.51)      yaná-t'ē  
               go (IRR)-1SG (FUT)  
               *I will go*

(3.52)      yān-t'eʔ  
               go (REALIS)-1SG (NONFUT)  
               *I went/am going/am about to go*

(Chung & Timberlake 1985:204-5)

In English, the intrinsic correlation between future time reference and the irrealis domain is reflected in the use of modal markers and progressive (imperfective) aspect in the expression of future time (see below).

Additionally, the various basic domains can be characterised in terms of dimensionality (Langacker 1983:157-158); time is one-dimensional, in contrast with, for example, space which can be two- or three-dimensional. The irrealis domain will be characterised as two-dimensional as it denotes varying degrees of likelihood of occurrence of events which may be either potential or hypothetical.

### 3.2.3 Summary

In this section, three theoretical constructs have been introduced: situation representations (SRs), operators and domains. These have been defined as follows:

#### *Situation representations*

A SR is a sub-propositional conceptual representation of a situation or event and is determined relative to the operator with immediate scope over it. A SR consists of the

propositional content of an utterance, that is, the conceptual information encoded in that utterance, inferentially enriched in accordance with the principle of relevance and any operators also within the scope of the operator relative to which the SR is determined.

### *Operators*

An operator is any expression which encodes procedural information constraining the interpretation of a SR.

### *Domains*

A domain is an element of the utterance interpretation context relative to which a SR is manipulated. Domains specify whether a SR is to be interpreted as a representation of an unverified event (the irrealis domain), as a representation of a past, present or future event (the temporal domain), and the manner in which an event occurs (various aspectual domains). The irrealis and temporal domains are basic, being irreducible to any other domain, whereas future time reference is an abstract domain, being defined in terms of both the temporal and the irrealis domains.

In the following section, the relation between SRs, operators and domains will be described with respect to the use of temporal and aspectual expressions in discourse. Modal auxiliaries are discussed in chapter 4.

## **3.3 Tense, Aspect and Time**

### 3.3.0 Introduction

I shall continue the discussion of MTA markers as exponents of procedural encoding at phase two, first by addressing the role of inference in temporal reference (3.3.1) and then that of tense (3.3.2) and aspect, in particular unmarked and progressive aspect in the expression of future time reference in English (3.3.3). In each case it will be shown that any account of the semantics of tense and aspect markers which views these as fully determining the truth-conditional content of their associated utterances is untenable (as is the case with pronouns). An alternative account, in which tense and aspect markers are analysed as encoding procedural information, will be proposed.

### 3.3.1 Temporal reference

Just as most pronominal expressions underdetermine their intended referents, so do tenses (in all the languages for which evidence is available) significantly

underdetermine the location of situations and events in time.<sup>43</sup> Even when a tense or aspect marker is used, pragmatic considerations (that is, assumptions derivable from the utterance interpretation context) still play a substantial role in determining the intended temporal reference of a situation or event. For instance, the temporal reference of the events in the following examples is determined largely as a result of pragmatic considerations or temporal adverbials (such as *just* in example (3.56)) rather than by the past or the 'present perfect' alone:

(3.53) A: Did you remember to lock the door?

B: No.

B: When?

(3.54) A: Have you eaten?

B: No.

B: ? Yes, frequently.

On a traditional Reichenbachian, truth-conditional account, (3.53) and (3.54) simply locate the events (E) [You remember to lock the door] and [You eat] respectively at a time prior to the temporal reference point (R), which in the absence of contextual information to the contrary is assumed to be the moment of speaking (S). Both (3.53) and (3.54) can be represented on this account by the following model:

Figure 10: a time-line analysis of the role of the past/present perfect



B's response "No" to both questions is true only if E did not occur at some point in the past. The most relevant interpretation of the questions in (3.53) and (3.54), however, is the one in which the time interval expressed through the past or present perfect has present relevance, hence B's response "No" to both questions would not ordinarily be taken to be false, even if B had in fact forgotten to lock the door a month previously. Evidence that the relevant time interval is inferred rather than encoded by the past or the present perfect comes from B's alternative responses: "When?" in (3.53), which indicates that the precise time reference was not encoded in A's utterance, and "Yes, frequently" in (3.54), which would ordinarily be taken to be literally true but infelicitous. (As Father Brown put it, "All language is used like that; you never get a

<sup>43</sup> Some languages, including Chinese, do not have tense markers (Janet Chan 22/5/96). In Chinese, temporal adverbials may, but need not, be used; in the absence of temporal adverbials, temporal reference, like pronominal reference with pro-drop in Chinese, is pragmatically determined.



question answered literally, even when you get it answered truly." G. K. Chesterton *The Invisible Man*) In (3.55), on the other hand the time interval between the occurrence of the event and the moment of speaking would ordinarily be assumed to be greater than that in (3.54). Being pragmatically determined, this can be overridden, for example by the addition of the temporal adverbial *just* which is truth conditional, as (3.56) illustrates. (For further discussion cf. Smith 1990, 1993 and Moeschler 1994).

(3.55) Mary has climbed Everest.

(3.56) ? Mary has just climbed Everest; she did it ten years ago.

Such evidence demonstrates that tense and aspect do not fully determine the temporal truth-conditions of the propositions of which they form part. This follows from the fact that tense and aspect are semantically underdetermined, hence there are "pragmatic aspects of what is said" in the expression of temporal reference, just as there are in the use of *and* for clausal conjunction.<sup>44</sup> (This semantic underdetermination of tenses has led Moeschler (1994) to view tense as non-truth-conditional which clearly contradicts "the standard semantic practice of calling a construction truth-conditional if and only if it contributes to the proposition expressed" (W&S 1993a:6) since tenses contribute to propositions even if they do not fully determine their temporal truth-conditions. A compromise position is adopted by Dahl (1985:13-14) who views truth conditions as partly accounting for the semantics of tense and aspect, which are primarily governed by conditions of use. This is the view towards which I shall tend; semantically tense and aspect are truth-conditional in the sense of *contributing* to truth-values (hence they function at phase two of utterance interpretation: the recovery of propositional form), but it is primarily pragmatic considerations or explicit temporal adverbials which *determine* the use and interpretation of tenses in discourse and the truth-value of an associated proposition.)

We have seen that inference plays a significant role in the determination of temporal reference. Temporal reference is not always wholly determined by inference, however; the inferential process can be constrained linguistically by the use of tense and aspect markers. In other words, tense and aspect markers encode procedural information constraining the inferential process of utterance interpretation with

---

<sup>44</sup> In §1.3.3 it was shown that the causal and temporal relations between clauses conjoined by *and* are pragmatically determined (since not encoded by *and*) but truth-conditional nonetheless:

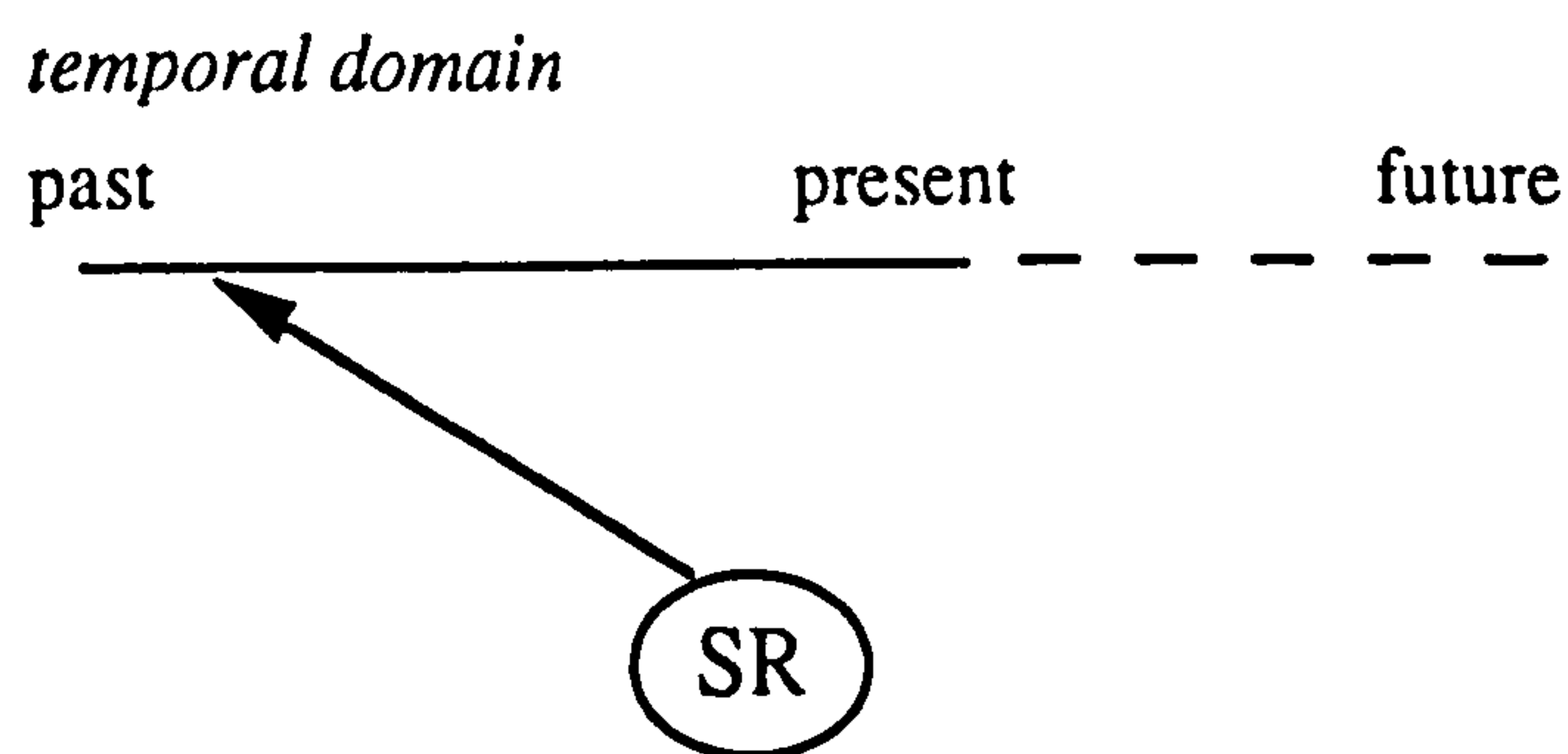
(1.12) a. What happened was not that Peter left and Mary got angry,  
but that Mary got angry and Peter left.

respect to the location of situations and events in time. Below, I shall briefly outline a procedural account of tense and aspect.

### 3.3.2 Tense

In this section, an account of temporal reference will be proposed in which tense markers are viewed as semantically underdetermined operators which manipulate any SR within their scope relative to the temporal cognitive domain, with the result that the event or situation described comes to be represented as occurring at some moment or period in time. Figure 11 below represents the function of a past tense marker, such as the English *-ed* suffix:

Figure 11: the function of the past tense<sup>45</sup>



However, a SR does not always require a tense marker in order to be related in a particular way to the temporal domain; this can be achieved by inference alone. This is what happens when non-past or unspecified tense is used in English: the addressee simply uses the most accessible context and the principle of relevance to determine the intended event time.<sup>46</sup> The English MTA system exploits this principle by assuming as part of a default context the most accessible temporal reference compatible with Optimal Relevance:

(3.57) I work in the city, so I have to leave the house pretty early. This morning, I go out to the garage as usual but the car isn't there!

The unspecified tense forms of 'work' and 'leave' are interpreted as habitual occurrences, given what we know about jobs and daily routines, but the unspecified

<sup>45</sup> The time line is represented by dashes in the future to illustrate that the future is necessarily irrealis.

<sup>46</sup> The use of non-past tense forms to refer to past time events is, however, restricted to certain highly constrained contexts, such as narratives (3.57) and jokes (3.58), and its use in past time contexts is marked. This is because past tense marking in English can be in 'harmony' with the context in which it is used, with the result that it can be semantically redundant and unmarked when used in conjunction with past temporal adverbials (see §5.1.2). For this reason I shall continue to talk of 'non-past' or 'unspecified' tense.

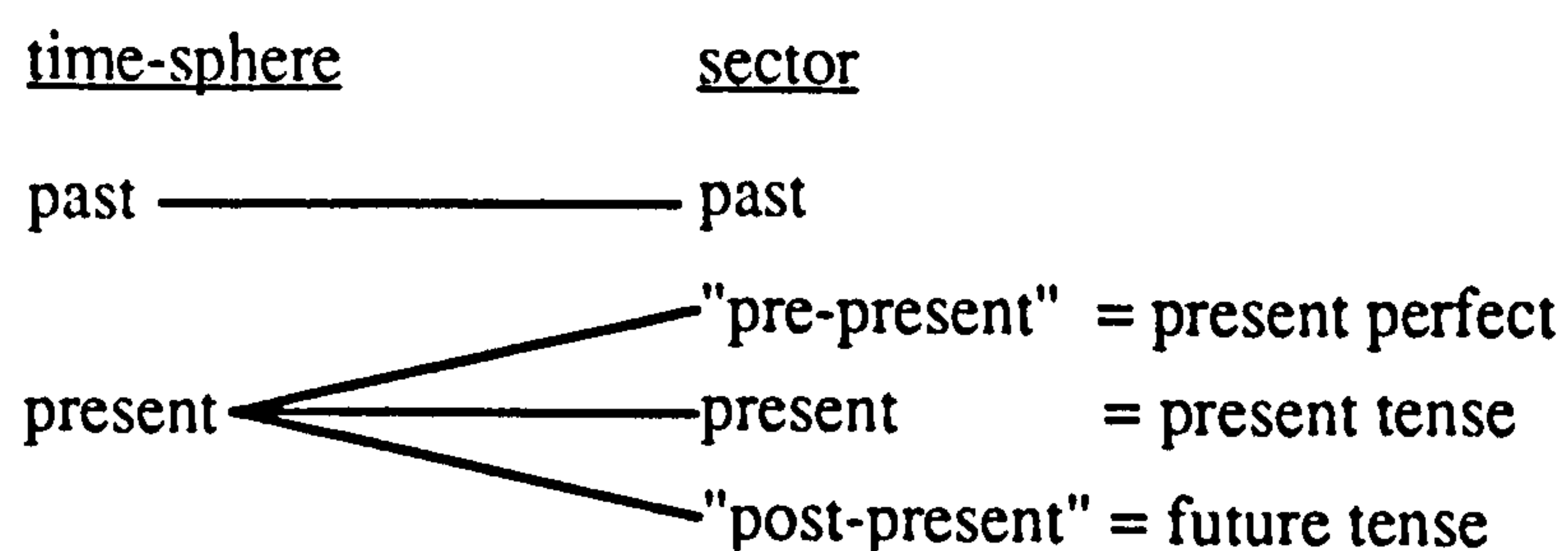
tense forms in the second sentence assume the temporal context set up by "This morning" and receive a definite past time interpretation.

Morphological tense in English takes the form of a binary opposition between past and unspecified (or non-past) tense constructions. The use of an unspecified tense construction indicates, as we have seen, that the addressee is to relate the situation or event in question to the currently most accessible temporal context, whereas the use of a past tense construction overrides this default assumption in favour of a past time interpretation (that is, the use of a past tense construction provides an explicit instruction to an addressee to relate a SR to the past time sector of the temporal cognitive domain). Further refinements arise from the use of modal auxiliaries, temporal adverbials, or perfect and/or progressive aspect in opposition to the simple or unspecified aspect.

Below, this characterisation of tense and temporal reference will be compared with a more traditional study of tense (Declerck 1989).<sup>47</sup> In the absence of a sufficiently powerful pragmatic theory, Declerck overstates the role of tense in determining temporal reference and understates the role of inference. The discussion will also incorporate aspects of an alternative relevance theoretic model of temporal reference (Moeschler 1994).

Declerck (1989) claims that English divides time into two 'time-spheres': past and non-past, corresponding to the two morphologically marked tenses in English; the non-past time-sphere includes  $t_0$  and the past time-sphere excludes  $t_0$  (where  $t_0$  is the zero-point or temporal reference point which is usually, but need not necessarily be, the time of utterance). Declerck divides the non-past time-sphere into three 'sectors' relative to  $t_0$ , giving four 'absolute sectors' in English. Each of these sectors corresponds to an absolute tense (Comrie 1985), that is, a tense in a direct rather than an indirect relation to  $t_0$ , as represented in figure 12, based on Declerck (1989:11-14):

Figure 12: Declerck's (1989) time-spheres and sectors



<sup>47</sup> See also Declerck (1991) for essentially the same analysis.

Declerck then distinguishes between literal and metaphorical uses of the four absolute tenses: when used literally, absolute tenses express time within their corresponding sector; however,

"Apart from this basic use, the absolute tenses are also used to represent a situation as if it belonged to a particular sector. In this use (in which there is a shift of temporal perspective to a sector that is different from the one to which the situation actually belongs) there is only a metaphorical relation between tense and time."

(Declerck 1989:83)

Two main objections to Declerck's model can be raised within the relevance theoretic framework: First, Declerck's claim that each of the three non-past time sectors corresponds to an absolute tense which can be used literally or metaphorically is incorrect. To begin with, whilst accepting his past versus non-past distinction, Declerck's inclusion in the four absolute sectors of a future *tense* (i.e. *will*) is unfounded given the evidence of §6.1. More importantly, even disregarding the claim made in §1.4.4 that it is impossible to uphold a *psychological* distinction between literal and metaphorical uses of language (cf. especially Gibbs 1989), it is incorrect to talk of literal and metaphorical uses of the present (i.e. non-past) tense. The non-past or unspecified tense construction simply licenses the addressee to relate a situation or event to the most accessible sector of the temporal domain. The fact that the most accessible temporal reference point is not always the present (i.e. the moment of speaking) does not entail a metaphorical usage. For example, the use of the unspecified tense construction in the 'historic present' requires that the most accessible context excludes present time considerations:

(3.58)            Here's a joke - a man walks into a pub and says...

(See also the examples in §3.3.3 below of the so-called 'present tense' with simple and progressive aspect used for future time reference.)

The unspecified tense construction also occurs in the 'present perfect', in the *have* auxiliary with morphologically non-past tense marking, in conjunction with the past participle form of the following verb. Here again the use of the unspecified tense indicates that the most accessible context is sufficient for optimally relevant utterance interpretation; however, the event in question is taken to have occurred prior to the reference time ( $E < R$ ) owing to the past participle. Thus in (3.54) and (3.55) repeated below, although the events described occurred in the past, to achieve optimal relevance the utterances must be processed against the most accessible temporal assumptions at the moment of speaking; a restricted period of recent time in (3.54) but a potentially much longer prior period in (3.55).

- (3.54) Have you eaten?  
 (3.55) Mary has climbed Everest.

I now turn to the second objection to Declerck's model. Moeschler (1994) has argued from a relevance theoretic perspective against the assumption (central to Lyons 1977, Comrie 1985 and Declerck 1989) that tense is intrinsically a deictic category. Moeschler (1994), on the basis of a comparison of the simple past and imperfect tenses in French, suggests that "it seems preferable to talk of anaphoric or deictic use of a tense rather than of intrinsically deictic or anaphoric tenses" (Moeschler 1994:88, my translation). For example, in (3.59) [Moeschler 1994 (97)] the simple past is being used deictically whilst the imperfect is being used anaphorically, whereas in (3.60) [Moeschler 1994 (98)] the roles are reversed:

- (3.59) Lorsque Marie entra, Jean téléphonait.  
*When Marie came in, Jean was on the phone.*
- (3.60) Alors que les joueurs discutaient avec l'arbitre, une bagarre éclata entre spectateurs.  
*Whilst the players were arguing with the umpire, a fight broke out in the crowd.*

In (3.59) the temporal reference point, R, is the deictic centre,  $t_0$ , and the event E1, 'Marie entra' (in the simple past), is located prior to R. The event E2, 'Jean téléphonait' (in the imperfect), is anaphoric in that it is determined by reference to E1, such that E2 contains E1. (3.59) can therefore be represented schematically as follows:

$$(3.59') \quad R = t_0; E1 < R; E1 \supset E2$$

In (3.60) the temporal reference point, R, is again the deictic centre,  $t_0$ , and the event E1, 'les joueurs discutaient' (in the imperfect), is located prior to R. The event E2, 'une bagarre éclata entre spectateurs' (in the simple past), is anaphoric being determined by reference to E1, such that E2 is contained in E1. (3.60) can also be represented schematically as follows:

$$(3.60') \quad R = t_0; E1 < R; E1 \subset E2$$

Moeschler (1994:93) concludes that the referential functions of tenses are deictic by default rather than being intrinsically deictic, with the result that in principle any temporal indicator (temporal adverb, date etc.) can function as a reference point and that only in the absence of such indicators is a deictic interpretation, i.e. an interpretation relative to the time of utterance ( $t_0$ ), necessary. This gives rise to the

following "interpretation procedure for temporal indicators" (paraphrased from Moeschler 1994:94):

Look for a temporal indicator from which to determine the intended temporal reference; if this fails, use the utterance interpretation context.

This 'interpretation procedure' is followed because explicit temporal indicators are more salient than the default context, which need not be the moment of speaking. Although derived from analysis of past tenses in French, Moeschler's interpretation procedure can also be applied to the use of the so-called simple present and present progressive in English with future time reference.

### 3.3.3 Aspect and future time reference

The following empirical analysis demonstrates that not only does tense significantly underdetermine temporal reference, but that there are constraints on the use of the unmarked and progressive aspect in English ('simple present' and 'present progressive') for future time reference which, were they semantically determined, would give rise to far too rich an ontology. The analysis below demonstrates that, given the appropriate relevance theoretic framework, there is no need for such a rich ontology, since these constraints arise from the interaction of pragmatic factors with the semantics of the two aspectual forms.

The semantics of the English aspectual system is less determinate than has been assumed. For example, Hatav (1993) characterises the English aspectual system in purely truth-conditional terms as a function from the intervals between reference time (R) and event time (E) to truth values, as follows: in simple aspect  $R \subset E$  (reference time contains event time), in progressive aspect  $R \supset E$  (reference time is contained within event time), and in the perfect  $R > E$  (event time occurs prior to reference time). This, however, makes the wrong prediction for the interpretation of the progressive with future time reference in (3.61a) (the event time does not contain the reference time) and wrongly predicts that (3.61a) and (3.61b) are truth-conditionally distinct:

- (3.61) a. She is leaving tomorrow.  
b. She leaves tomorrow.

Whilst (3.61a) and (3.61b) are not truth-conditionally distinct, there are constraints on the use of unmarked and progressive aspect with future time reference which

condition their distribution and interpretation. These will be discussed below, first in relation to unmarked aspect and then in relation to the progressive aspect.

The use of the 'simple present' form of a main verb for future time reference is conditioned by three constraints:

i) With the exception of stylised usage such as that of weather forecasts, the simple present can only be used with reference to events that are planned or absolutely certain, hence the unacceptability of (3.62) and (3.63) which can be neither planned nor absolutely certain:

(3.62) ? It rains in a minute.

(3.63) ? You get better tomorrow.

ii) It must co-occur with an explicit future time adverbial, otherwise the default context (often present time) is assumed:

(3.64) a. I work nights as of next week.

b. I work nights. (obligatory habitual interpretation)

iii) The situation referred to is one over which the subject (not, as Wekker 1976:82 claims, the speaker) has no control:

(3.65) Mary has a birthday next week.

(3.66) ? Mary has a party next week.

(3.67) It's the best news I've had since midsummer. Tomorrow we celebrate!

(3.68) I have done my business, Sam. We return to Lyme tomorrow.

(3.69) Now go to the house, all of you. I have phone calls to make, work to do. Tomorrow we leave.<sup>48</sup>

Although (3.60) to (3.62) appear to contradict this constraint, the situations referred to are ones over which the speaker alone, rather than the subject (inclusive *we*, that is addressees plus speaker), has control; that is, they are arrangements imposed on the addressees by the speaker. In this connection, Klinge (1993:326) notes that a future situation or event can be brought about either by "the motivated production of an activity with the intention of bringing a consequent state about" (Klinge, p.c. 7/10/94)

---

<sup>48</sup> (3.67) Tolkein, J. R. R. 1977 *The Fellowship of the Ring* Toronto, Methuen (page reference lost); (3.68) Fowles J. 1973 *The French Lieutenant's Woman* New York, Signet p.256; (3.69) Bradbury, R. 1969 *Dandelion Wine* New York, Bantam p.183. Cited in (Hurtle & Curat 1986:56).

(what he calls an 'agent-event'), or by some situation in the world without any intentional activity (that is, a 'world-event'). I would suggest that the simple present with future time reference cannot be used to represent agent-events in which the agent and the subject of the clause are coreferential.

The use of the 'present progressive' form of a main verb for future time reference is similarly conditioned by three constraints:

i) similar to the simple form, the progressive can only be used with events which are planned, hence the unacceptability of (3.70) and (3.71) but not of (3.72) or (3.73) below:

- (3.70)           ? It is raining in a minute.
- (3.71)           ? You are getting better tomorrow.
- (3.72)           Are you giving a paper? (said over lunch at the start of a conference)
- (3.73)           Mary's having a party next week.

ii) it must either co-occur with an explicit future time adverbial, or the utterance interpretation context must preclude a present time interpretation (i.e. the most accessible temporal reference point must not be the present):

- (3.72)           Are you giving a paper? (said over lunch at the start of a conference)
- (3.73)           Mary's having a party next week.

iii) "The future use of the present progressive is limited to actions brought about by human endeavour." (Quirk *et al* 1985:215) In other words, the progressive can not be used to represent a future world-event, that is an event brought about by some situation in the world without any intentional activity :

- (3.46)           ? Mary's having a birthday next week.

Overriding this is the restriction on the use of the progressive with verbs of state or achievement (Vendler 1967: Chapter 4) which have no continuous aspect:

- (3.47)           ? I'm being here all day tomorrow.

The third constraint on both forms (that the 'simple present' cannot be used to represent agent-events where the agent and the subject are coreferential, and that the 'present progressive' cannot be used to represent world-events) is applicable only with



future time reference. The principle of relevance accounts for this in that the additional cognitive effort necessitated by the effort needed to choose (for the speaker) between the various future expressions in English and to decide (for the addressee) why one form was used in preference to another, is offset by additional contextual effects. The additional contextual effects available with present time uses of the two forms (habitual occurrence versus current and ongoing occurrence, etc.) are not applicable to future time reference, so different contextual effects arise. These effects are illustrated in the following minimal pairs:

(3.48) a. I have an extra day off next week.

(3.49) a. I'm having an extra day off next week.

The simple present implies an arrangement imposed on the subject, whereas the present progressive implies an arrangement made by the subject:

(3.48) b. I have an extra day off next week because the office is closed.

(3.49) b. I'm having an extra day off next week to go to a friend's wedding.

In summary, the first constraint on the use of both the 'simple present' and the 'present progressive' for future time reference falls under Moeschler's interpretation procedure for temporal indicators. In the case of a representation of an event using the 'simple present' form of a verb, the time reference may be universal ('Oil floats on water'), habitual ('I cycle to work'), etc. depending on the nature of the event and any accompanying adverbials. For this reason, a future time temporal adverbial must be actualised or understood for the 'simple present' to refer to the future.

The progressive, on the other hand, is typically used to represent events which are ongoing and actual; when such an interpretation is ruled out by the context/cotext, future time reference may be inferred even in the absence of a future time adverbial. This is due not to progressivity *per se* so much as to the fact that the 'progressive' form in English expresses imperfectivity (that an event is not yet completed or finished). The fact that, in its future time uses, the 'present progressive' does not always represent situations or events as ongoing, is evidence for treating the 'progressive' as semantically an imperfective marker. In the spirit of Modified Occam's razor, there is no need to posit a further progressive sense.

If the 'present progressive' is viewed as an imperfective marker, its future time uses come as no surprise. Bybee *et al* (1991), in a cross-linguistic study, identify three major types of future, one of which is a future based on aspectual forms "whose main

function is to mark imperfective, and, in the rare case, perfective, but which can also be used for statements about the future." (1991:18) They continue:

"For both groups, we suspect that the future reading is only obtained when the context warrants it; that is, the semantics of the gram<sup>49</sup> does not explicitly contain the features 'prediction' or 'future time reference', but is much more general, and allows future interpretation if it is supported by the context."

(Bybee *et al* 1991:21)

As we have seen, this is the case for English; furthermore, I suggest that the reason why a future time interpretation is possible in the absence of a temporal adverbial only with the imperfective (progressive) form lies in the (inferential) relation between imperfective aspect and the irrealis domain, which, as we saw in §3.2.2 is one of the basic cognitive domains to which the abstract future domain reduces. This is demonstrated through the "imperfective paradox" in which "the problem is to give an account of how [John was drawing a circle] entails that John was engaged in a bringing-a-circle-into-existence activity but does not entail that he brought a circle into existence" (Dowty 1977:46). The ontology of the imperfective does appear to require the additional stipulation that for future time reference, the situation or event referred to must be the result of an agent-event, not a world-event, given (3.46) and (3.73):

(3.46)           ? Mary's having a birthday next week.

(3.73)           Mary's having a party next week.

Inferences to the irrealis domain can also be made on the basis of the use of modal expressions, and in chapter 4 I shall argue that attempts to characterise the semantics of the English modal auxiliaries in model theoretic terms involving possible worlds and modal logic are unsatisfactory. Before discussing modal auxiliaries, however, I shall summarise the findings of the current chapter and discuss the prospects for a procedural account of modal expressions.

---

<sup>49</sup> An abbreviation of 'grammatical marker': "Formally, grammatical morphemes may be affixes, stem changes, reduplication, auxiliaries, particles, or complex constructions such as English *be going to*. We refer to all of these types equally as grammatical morphemes and for convenience shorten this term to "gram"." (Bybee *et al* 1994:2)

## 3.4 Summary and Prospects

### 3.4.0 Introduction

In §3.1 and §3.3 I demonstrated that pronouns and tense and aspect markers are not directly truth-conditionally interpreted. It was shown that the truth-conditions of utterances containing either nominal, temporal or aspectual indexical expressions are to a large extent inferentially determined. The function of pronouns, and tense and aspect markers is simply to constrain the inferential processes leading to the identification of propositional form. Pronouns, tense markers and aspect markers therefore encode procedural information manipulating (that is, constraining the inferential interpretation of) sub-propositional conceptual representations. I shall briefly review some of the arguments which led to these conclusions.

### 3.4.1 Summary

#### *Pronouns*

In the case of pronouns, we saw that reference assignment is a largely inferential process, as illustrated in (3.1) and (3.2), but that there are restrictions on the discourse entities with which pronouns can be coindexed, as (3.3) illustrates:

- (3.1) Jane<sub>i</sub> hugged Mary<sub>j</sub> because she<sub>i</sub> was happy.
- (3.2) Jane<sub>i</sub> hugged Mary<sub>j</sub> because she<sub>j</sub> was crying.
- (3.3) Jane<sub>i</sub> hugged Mary<sub>j</sub> and then she<sub>i/\*j/\*k</sub> hugged Harry.

I concluded that pronouns encode procedural information concerning the level of relative accessibility of (mental representations of) their intended referents, in line with Accessibility Theory. According to the Accessibility hierarchy (Ariel 1991:449) pronouns are markers of relatively high accessibility and will therefore be used only when a speaker judges that the mental representation of the entity to which it refers is highly accessible to the addressee, relative to alternative mental representations.

This was illustrated with reference to Kiswahili, in which bound pronominals (the SP and OP) are markers of higher accessibility than emphatic pronouns and full NPs; this enables them not only to function referentially in the absence of an overt NP but also to topicalise a coreferential NP when one is present.

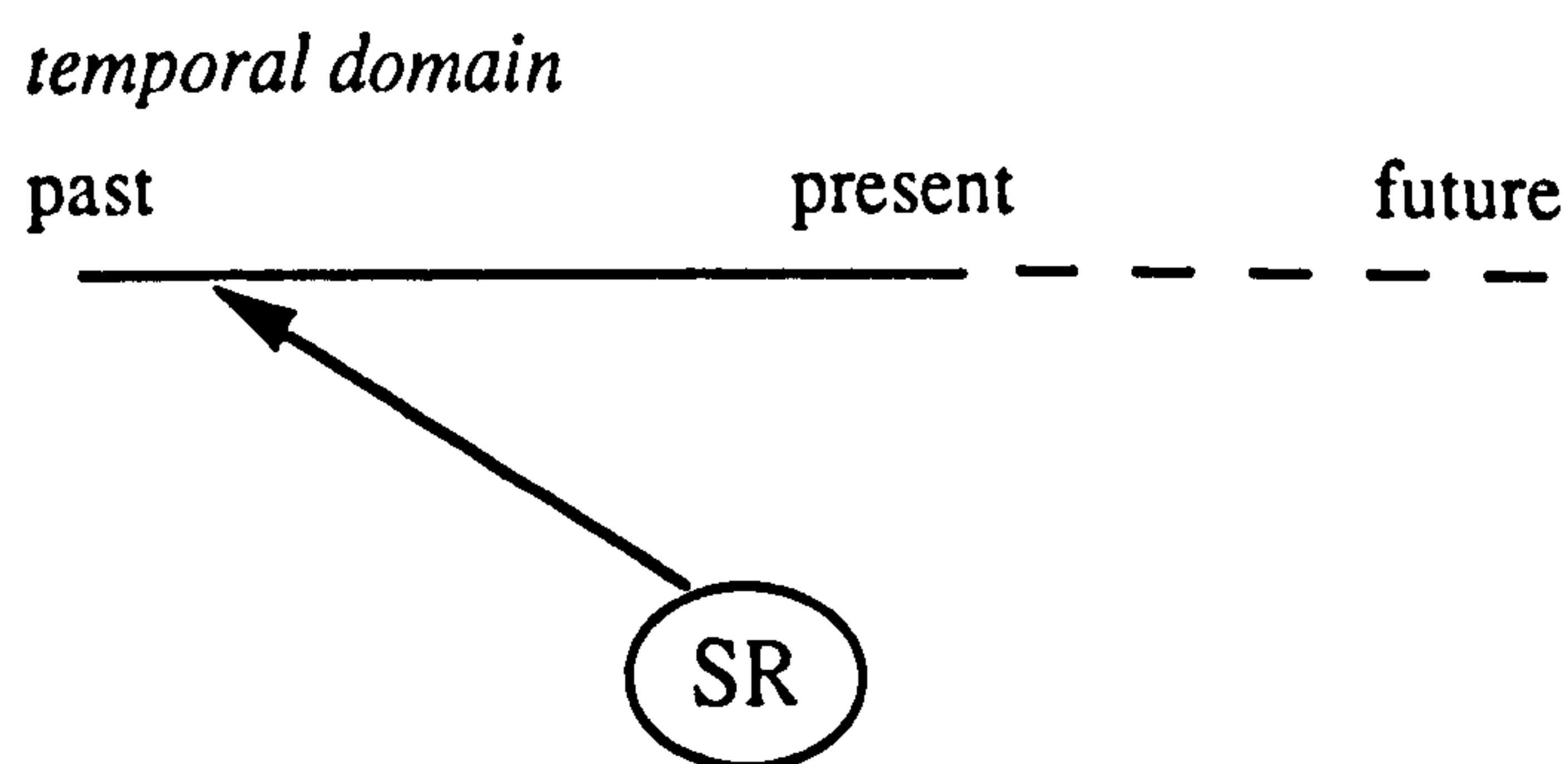
### *Tense markers*

Inference plays an important role in determining the intended temporal reference of a situation or event, as examples (3.53), (3.54) and (3.55) demonstrate. In (3.53) and (3.54), unlike (3.55), it is not enough simply to locate the events described at some unspecified moment in the past to recover the intended proposition:

- (3.53) Did you remember to lock the door?  
(3.54) Have you eaten?  
(3.55) Mary has climbed Everest.

The role of tense markers is to constrain the inferential processes involved in the determination of temporal reference. More precisely, tense markers were characterised semantically as operators which manipulate any SR within their scope relative to the temporal cognitive domain, as represented with respect to the past tense in figure 11:

Figure 11: the function of the past tense



### *Aspect markers*

I demonstrated that a directly truth-conditional account of aspect, such as that proposed by Hatav (1993) which analyses simple aspect as encoding  $R \subset E$  (reference time contains event time), and progressive aspect as encoding  $R \supset E$  (reference time is contained within event time) wrongly predicts that (3.61a) and (3.61b) are truth-conditionally distinct:

- (3.61) a. She is leaving tomorrow.  
b. She leaves tomorrow.

The use of unmarked aspect for future time reference requires an explicit future time adverbial, but imposes constraints on the event in question (it must be certain and must be an event over which the subject of the clause has no control). The use of the imperfective aspect for future time reference requires that the utterance interpretation context preclude a present time interpretation, but the use of the imperfective aspect

also imposes constraints on the event (it must be planned and its occurrence be due to an agent-event in which the subject of the clause is the agent). These constraints are not truth-conditional, since (3.61a) and (3.61b) are not truth-conditionally distinct, nonetheless they affect the distribution and interpretation of futurate situations and events expressed through unmarked or imperfective aspect.

### 3.4.2 Prospects: Modality

The current chapter has demonstrated, with respect to pronouns, and markers of tense and aspect, that the conceptual/procedural distinction can be usefully extended to phase two of utterance interpretation. In chapter 4 I shall illustrate this in greater detail with respect to the English modal auxiliaries, and in chapter 5 elaborate the conceptual/procedural distinction at phase two to accommodate semantic change and variation.

In chapter 4, I will demonstrate that, like pronominal referring expressions and tense and aspect markers, modal expressions also underdetermine the truth-conditions of propositions derived from utterances in which they occur. Given the analyses proposed above of pronouns, tense markers and aspect markers as exponents of procedural encoding, we would expect a procedural account to offer some prospect of providing a satisfactory account of the modal auxiliaries as well. Such an account will be proposed for *can*, *may*, *must* and *should*, drawing primarily on Groefsema (1995a). These English modal auxiliaries will be discussed in some detail, in part owing to the complexity of the functions and distributions of the various modal auxiliaries, and also to provide a link to the more detailed discussion of grammaticization and the expression of future time reference in English in chapters 5 and 6.

## 4 THE ENGLISH MODAL AUXILIARIES

I shall begin this chapter by discussing the inadequacies first of purely truth-conditional accounts of the English modal auxiliaries based on the application of modal logic to the characterisation of linguistic expressions of modality, and then of traditional descriptive accounts, both polysemous and monosemous. I shall then consider and reject two relevance theoretic accounts of the modal auxiliaries, and conclude by developing an alternative relevance theoretic model based on a procedural analysis of *can*, *may*, *must* and *should*.

### 4.1 Theoretical Preliminaries

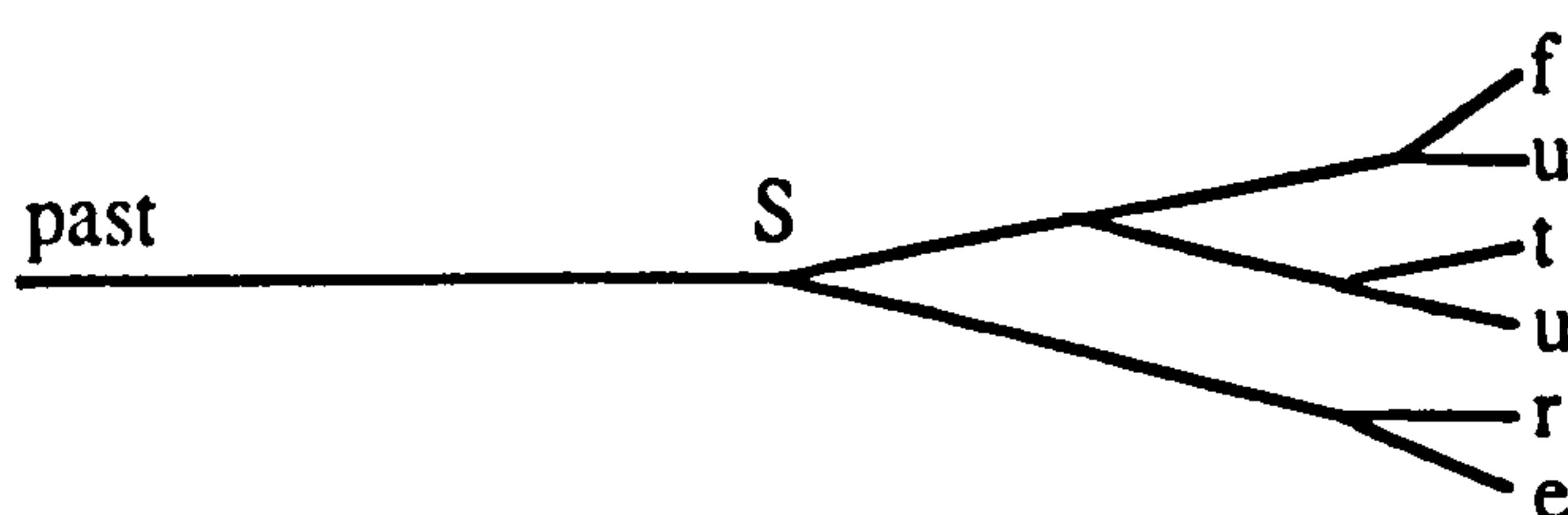
#### 4.1.0 Introduction

In this section I shall demonstrate that attempts to characterise the English modal auxiliaries in directly truth-conditional terms fail to provide satisfactory accounts of the interpretation of utterances in which they occur. I shall begin by rejecting the possible worlds metaphor as a satisfactory basis for a semantic characterisation of any of the modals, and will then also reject traditional descriptive accounts of the modals which are not based explicitly on the notion of possible worlds. I shall conclude that only a semantic/pragmatic theory which takes into account the semantically underdetermined nature of the modals and allows for pragmatically determined aspects of truth-conditional content can hope to provide an adequate characterisation of the English modal auxiliaries.

#### 4.1.1 Logical analyses of the modals

Although both the temporal and the irrealis domains are cognitive, they have been represented in model theoretic terms on the basis of possible worlds. For example, the branching futures model (Thomason 1970, Dowty 1977, Tedeschi 1981) views time as branching out into a number of alternative worlds subsequent to the moment of utterance (S); this is represented schematically in Figure 13 (taken from Matthews 1989:190, based on Dowty 1977:63):

Figure 13: the branching futures model of time



The possible worlds in the branching futures model are located temporally at a time subsequent to the moment of speaking, but this model can be adapted to take into account possible worlds existing simultaneously with the actual world by making reference to alternative worlds posited at some prior moment. An alternative model is the 'parallel worlds' model devised by Matthews (1989:197ff.). According to this model, a speaker at time *S* may entertain a number of parallel 'intensional' worlds (the realis world plus a number of irrealis - or for Matthews 'potentialis' and 'irrealis' - worlds), for which the temporal specifications of events within these worlds are variable. Unlike the branching futures model, in which possible worlds diverge from the actual world at some point on the time line, the parallel worlds model "allows events, (whether factual or fictional) to be placed on or against the time line while viewing them, from the point-of-view of the speaker, as real, potential or unreal." (Matthews 1989:198)

Three basic approaches to possible worlds can be distinguished (Haack 1978:190-1): the linguistic approach which views possible worlds as maximally (syntactically or semantically) consistent sets of sentences (e.g. Hintikka 1969); the conceptualist approach which talks in terms of how we could conceive the actual world to be different (Kripke 1972); and the realist approach which views possible worlds as real abstract entities, independent of either language or thought (Lewis 1973 chapter 4). My argument is that any appeal to the notion of possible worlds, however characterised, provides an inadequate basis for a satisfactory analysis of the modals. I shall begin by outlining some of the philosophical problems encountered by the possible worlds metaphor and then demonstrate the inability of this metaphor to account for the behaviour of the modal auxiliaries.

The possible worlds metaphor is closely connected with modal logic. The first modal logic was developed (Lewis 1918), not to provide a formal treatment of modality, but in response to a dissatisfaction with the notion of material implication. The problem was that intuitively, for *p* to imply *q* it is not enough to say that it is not the case that *p* is true and *q* is false, but rather that *p* could not be true and *q* false. To capture this intuitive notion, strict implication was proposed, defined as follows:

$$A \prec B = \text{df } L(A \rightarrow B), \text{ where } LA = \text{df } \sim M \sim A$$

where L is a modal operator (to be read as 'in all possible worlds') which can be defined in terms of the other primitive modal operator M (to be read as 'in at least one possible world').<sup>50</sup>

One problem facing modal logic is that modal contexts are 'referentially opaque' (Quine 1961). For example, the use of 'the number of planets' to refer to '9' is not purely referential, that is, it does not refer directly in certain contexts. In conjunction with a propositional attitude, for instance, substitution of '9' for 'the number of planets' turns the true (4.1) into the false (4.2):

- (4.1) Peter believes that the number of planets is eight.  
 (4.2) Peter believes that 9 is eight.

Referential opacity afflicts modal contexts as well. Each of the modal systems developed is self-contained with the result that a given modal logic, say T, can neither interact with another modal logic, say S5, nor with a non-modal system. To combine two incompatible systems results in failure of substitutivity, as the following argument from Quine (1961) demonstrates. Given the two true premises (4.3) and (4.4), substitutivity on the basis of true identities results in the falsehood (4.5):

- (4.3) L (9 > 7)  
 (4.4) The number of planets = 9  
 (4.5) L (The number of planets > 7)

As a result of the referential opacity of modal contexts, we cannot quantify into them; that is, we cannot go from (4.3) to (4.6) by a process of existential generalisation:

- (4.6)  $(\exists x) L (x > 7)$

Attempts to overcome the problem of quantification into modal contexts (cf. Linsky (ed) 1971) have focused on the notion of 'essentialism', that is, the idea that objects have some of their properties necessarily, and others only contingently, akin to Fregean ordinary sense and reference (Frege 1892) respectively. The claim is that,

---

<sup>50</sup> However, strict implication still seems inappropriate as the formal counterpart of our intuitive notion of implication given the 'paradoxes' of strict implication:

$$Lp \rightarrow (q \prec p) \quad \text{and} \quad L\sim p \rightarrow (p \prec q)$$

"i.e. a necessary proposition is strictly implied by any proposition whatever, and an impossible proposition strictly implies any proposition whatever." (Haack 1987:197) Such considerations led to the development of 'relevance logics' (cf. Anderson & Belnap 1975) which remain distinct from relevance theory.



when quantifying into modal contexts, substitutions should only be made on the basis of essential properties, that is, only substitution on the basis of identity of sense is permissible. Quantification into modal contexts becomes permissible given some form of essentialism, since (4.6) can be paraphrased as "there is an  $x$  which, in all possible worlds, is greater than 7"; this entails that being greater than 7 is essential to an entity being the  $x$  of (4.6). Furthermore, only by stipulating that referring expressions have constant reference in all possible worlds can propositions such as (4.7) below make any sense, otherwise there is no constraint keeping the referent of a variable from changing as we pass from one possible world to another:

(4.7)  $(\exists x) M (x = \text{the number of planets})$

In short, some form of essentialism is necessary for quantification into modal contexts, but this places strict limits on what can count as a possible world, and hence strict limits on what counts as permissible quantification into modal contexts.

The failures of substitutivity noted by Quine, which derive from combining incompatible systems within a single argument, do not occur when sentences incorporating modal auxiliaries and non-modal sentences are combined in natural language, however. The fact that utterances containing modal auxiliaries are easily combined with non-modal utterances demonstrates that the modal auxiliaries do not express a self-contained logical system. Whilst (4.5) ( $L(\text{The number of planets} > 7)$ ) is ruled out by the fact that (4.4) ('The number of planets = 9') is not essential to the identity of the number of planets, (4.8) below, in which the ordinary language term *must* has replaced the logical operator  $L$ , is an acceptable conclusion given the same two premises:

(4.3)  $L (9 > 7)$

(4.4) The number of planets = 9

(4.8) The number of planets must be greater than 7.

(4.8) does not express (4.5), rather it expresses that, given the truth of the premises (4.3) and (4.4), the proposition that the number of planets is greater than seven can be deduced and is therefore true in the actual world. The use of *must* in (4.8) is a comment on the validity of the argument rather than on the necessity of its conclusion.

Despite these problems with modal logic and the mismatch between modal logic and its supposed natural language (English) counterpart, the system of modal auxiliaries,

the possible worlds model has continued to be invoked. To conclude this section, I shall discuss the problems encountered by a relevance theoretic account of the modals *may*, *can* and *must* (Papafragou 1995) based on the possible worlds metaphor.

Papafragou (1995:3) invokes the possible worlds model in the following way:

"A state of affairs represented by a given proposition  $p$  will be possible in view of the speaker's knowledge if and only if there is a world  $w_2$  which is accessible from  $w_1$  and the state of affairs holds in  $w_2$ . A state of affairs represented by a proposition  $p$  will be necessary in view of the speaker's knowledge if and only if the state of affairs holds in all worlds accessible from  $w_1$ ."

In addition to the constraints imposed on possible worlds in order to allow quantification into modal contexts, Papafragou (1995) places additional cognitive constraints on the possible worlds relative to which the semantic characterisations of the modals are determined. A possible world which is accessible from  $w_1$  (the actual world) must be compatible with the speaker's knowledge, which is defined as differing in terms of a minimal number of assumptions only. What constitutes a 'minimal' number of assumptions here is unspecified.

Aside from this problem, two important objections can be raised against Papafragou's account. First, while it is clear how *may* and *must* relate to the possible worlds model, it is not stated how *can* (let alone any of the other modals) relate to it. *May*, *can* and *must* are defined as follows (Papafragou 1995:3):

*May*: It is possible that  $p$ .

*Can*: It is potential that  $p$ .

*Must*: It is necessary that  $p$ , or: It is not possible that not- $p$ .<sup>51</sup>

In relation to the possible worlds model, *may*  $p$  expresses that there is at least one possible world accessible from  $w_1$  such that  $p$  holds in it, and *must*  $p$  expresses that  $p$  holds in all the possible worlds accessible from  $w_1$ . Since it is not stated how a situation in which  $p$  is potential (rather than possible or necessary) relates to any possible worlds, it is unclear how *can* is intended to be determined within the possible worlds model.

---

<sup>51</sup> Note that Papafragou's characterisation of *must* in terms of *may* (possible) exactly corresponds to the definition of the logical operator  $L$  in terms of  $M$ :  $LA = df \sim M\sim A$ .

The second objection to Papafragou's account is that it is not stated whether  $w_1$  itself is permissible as a possible world. In other words, is it possible that  $w_1 = w_2$ ? Given (4.9) it would seem that this must be possible:

(4.9) He may be a university professor, but he sure is dumb.

[Papafragou 1995 (24)]

This suggests that if  $p$  is necessary then it is true in  $w_1$  as well as in every other possible world accessible from  $w_1$ . In the case of analytic sentences (sentences which are true by definition) this is obviously the case:

(4.10) A bachelor must be unmarried.

(4.11) A bilabial nasal must be made with the lips closed.

In non-analytic sentences, however, *must* often reduces the strength of a speaker's commitment to  $p$  in comparison to simply stating  $p$  in a non-modal expression:

(4.12) *Knock, knock* "That must be the postman."

(4.13) *Knock, knock* "That's the postman."

Whilst (4.13) is felicitous in a situation in which the speaker can see the postman standing outside the front door holding a parcel, (4.12) is not. (4.12) suggests that the speaker has deduced that the postman has just knocked on the basis of incomplete evidence (e.g. it is early in the morning, the postman always knocks twice, it is not a Sunday, etc. but excluding direct visual evidence). Surely, on the basis of such evidence there is a possible world accessible from  $w_1$  in which it is not the postman knocking at the door, and hence to characterise *must p* as 'true in every possible world' is to make too strong a claim.

In conclusion, any attempt to characterise the semantics of the modal auxiliaries in terms of a possible worlds model faces considerable, if not insurmountable difficulties. Papafragou (1995) will be discussed further (in §4.2) in relation to the question of whether modal auxiliaries should be characterised as encoding conceptual or procedural information. To conclude this section, I shall discuss non-relevance theoretic analyses of the modals which do not make explicit reference to the possible worlds metaphor.

#### 4.1.2 Traditional descriptive analyses of the modals

Traditional accounts of the modal auxiliaries can be divided broadly into those which propose polysemous analyses (Leech 1971, 1987; Coates 1983; Palmer 1979, 1986; Quirk *et al* 1985; and Hermerén 1978 who identifies twenty distinct modal meanings for nine modal auxiliaries) and those which propose monosemous (unitary) analyses (Ehrman 1966; Lyons 1977; Kratzer 1977, 1981; Perkins 1982, 1983; Sweetser 1982).<sup>52</sup>

Walton (1991:326) observes that under the polysemy approach, the meanings of *can* are usually given as 'possibility', 'permission' and 'ability', corresponding neatly to the three categories of logical modality: epistemic - the modality of knowledge and belief, deontic - the modality of permission and obligation, and dynamic - the modality of ability and willingness. However, not all polysemous accounts distinguish between the various meanings or uses of the modals according to this three-way division (e.g. Palmer 1979, 1986, and Sweetser 1989). Problems arise when the interpretation of a modal in a given sentence cannot be paraphrased in terms of any one of the proposed meanings derived from the logical modalities. Walton (1991:332) argues that semantic considerations alone are incapable of determining whether an utterance of a sentence like (4.14):

(4.14) Freddie can recite a poem.

expresses 'possibility', 'permission' or 'ability', or even some combination of these (Freddie is willing and able to recite a poem, for example). If the modals were genuinely ambiguous between distinct meanings, we would expect to be able to disambiguate them, but this is clearly not the case.

The polysemy approach also encounters problems in accounting for uses of modals which are neither indeterminate between any of their proposed meanings, nor interpretable as a combination of any of these. Groefsema (1995a:57) suggests the following example:

(4.15) You must come to dinner sometime.

---

<sup>52</sup> An intermediate position is adopted by Sweetser (1989) in which the root (deontic and dynamic) meanings of the modals are taken as basic and are metaphorically mapped from the external to the mental world yielding epistemic interpretations. Groefsema (1995a:58-60) discusses Sweetser (1989) in some detail.

and observes that an utterance of this sentence would not normally be interpreted as (4.16) or as (4.17) but rather as (4.18), or something similar:

(4.16) It is necessary that you come to dinner sometime.

(4.17) You are obliged to come to dinner sometime.

(4.18) We would like you to come to dinner sometime.

The interpretation of utterances containing modals is therefore highly context dependent. In the spirit of Modified Occam's Razor, and in view of the fact that the various meanings attributed to the modals in polysemous analyses "invite an explanation which stresses their unity, since we intuitively link them" (Walton 1991:327), a monosemous (semantically unitary) approach is to be preferred.

The traditional monosemous analyses cited above suffer from the lack of an adequate pragmatic theory by which to account for the role played by context in utterance interpretation. More fundamentally, the role which is assumed for pragmatics is essentially Gricean, and can be summarised in Gazdar's equation (Gazdar 1979:2): pragmatics = meaning - truth-conditions. As the arguments presented in §1.3.3 demonstrated, this delimitation of pragmatics is untenable. The following analysis demonstrates this with respect to the modal *can*. The sentence in (4.14) is capable of being used to express more than one proposition:

(4.14) Freddie can recite a poem.

(4.19) Freddie has the ability to recite a poem.

(4.20) Freddie is permitted to recite a poem.

Given that *can* is not polysemous, the differences in propositional content between (4.19) and (4.20) must be pragmatic in origin. As a result, only a pragmatic theory which, like relevance theory, recognises pragmatic aspects of 'what is said' can hope to provide an account of how utterances of sentences like (4.14) are interpreted.

#### 4.1.3 Summary

We have seen that attempts to characterise the modals in a directly truth-conditional way are unsatisfactory, whether this is attempted through translation into a modal logic defined in terms of possible worlds, or through the development of a rich ontology for each of the modals. Attempts to translate modal auxiliaries into a modal logic defined in terms of possible worlds encounter problems, both internal to modal logic (referential opacity and the debate over essentialism), and empirically when

modal auxiliaries are characterised as the natural language counterparts of a logical system (even as regards the most obvious candidates for such an account, *may* and *must*). Alternative analyses of the modals as polysemous encounter problems over how to account for indeterminacy in interpretations of utterances containing modals; and previous monosemous accounts, in the absence of an appropriate pragmatic theory, failed to account for pragmatically determined aspects of truth-conditional interpretation.

Given the relevance theoretic analyses, provided in chapter 3, of tense and aspect markers (which also defy adequate description in purely truth-conditional terms), it is possible that the English modal auxiliaries might be adequately described within a procedurally based account. This analysis will be given in §4.2 in which two relevance theoretic accounts of modality (one purely procedural and one purely conceptual) will be rejected in favour of a procedural reformulation of Groefsema's (1995a) relevance theoretic analysis of *may*, *can*, *must* and *should*.

## **4.2 Can, May, Must and Should**

### **4.2.0 Introduction**

The aim of this section is to provide an explicit procedural account of the use and interpretation of a subset of the modal auxiliaries. For the time being, I shall not address the relation between procedural encoding at phase two of utterance interpretation and grammaticization; rather, I shall provide a purely synchronic procedural analysis of *can*, *may*, *must* and *should* taking account of the systematic semantic/pragmatic interdependence of these forms.<sup>53</sup> To this end, I shall briefly review two recent relevance theoretic accounts of the English modals: Klinge's (1993) procedural account of the modals as encoding potentiality, and Papafragou's (1995) conceptual account of the modals in which epistemic uses of *may* and *must* are viewed as contributing to higher-level explicatures. These will be rejected in favour of Groefsema's (1995a) analysis of *may*, *can*, *must* and *should* which, although agnostic concerning the conceptual/procedural distinction, will be reformulated in procedural terms. The primary purpose of this section is not to provide a comprehensive account of the semantics and pragmatics of the modals *may*, *can*, *must* and *should*, nor even to refine the relevance theoretic proposals discussed; my intention is simply to determine whether these modals encode procedural information,

---

<sup>53</sup> For historical accounts of the English modal auxiliaries see, for example, Lightfoot (1979, 1991) and Warner (1983, 1993).

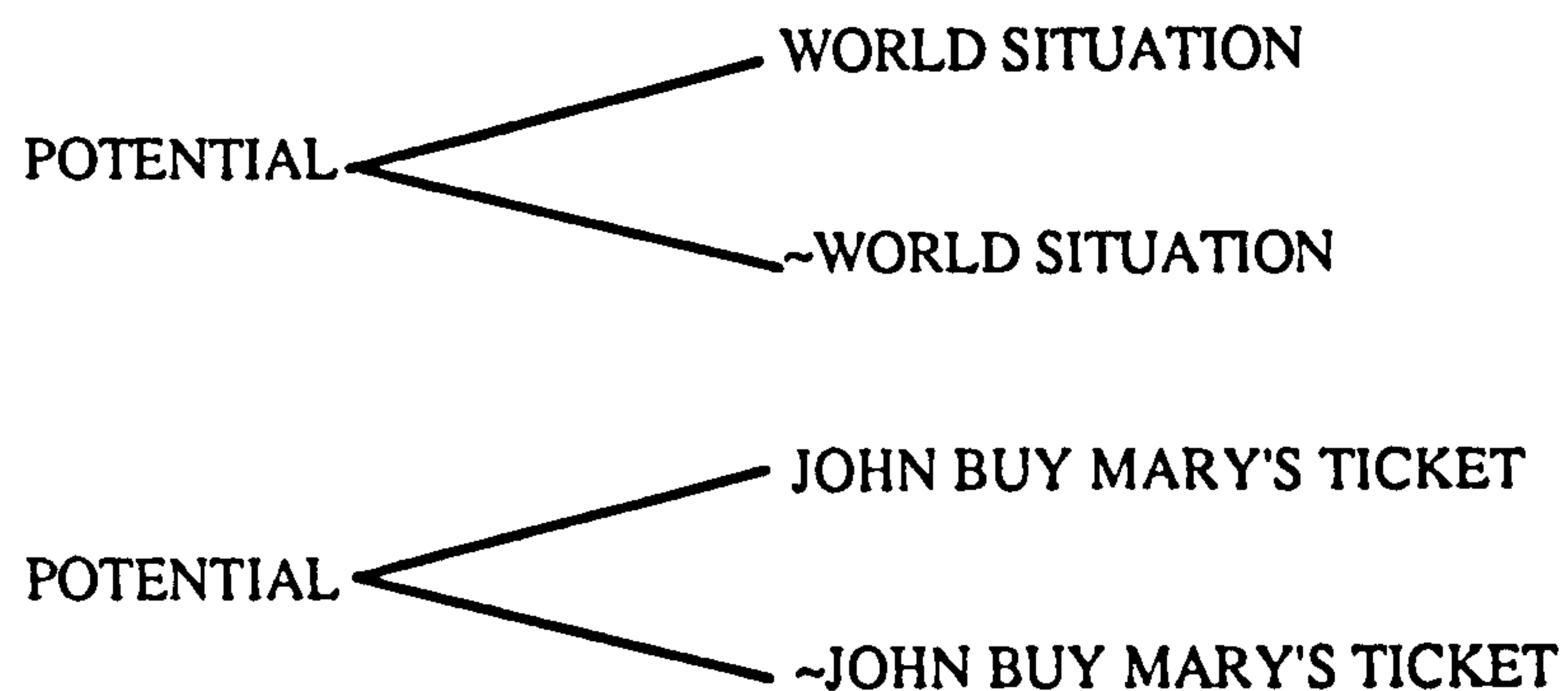
and if so how they constrain the interpretation of situations and events described in utterances containing them.

#### 4.2.1 Modals as procedural markers of potentiality: (Klinge 1993)

Klinge (1993) provides an account of the 'central' modals - *can, may, will, shall* and *must* - in purely procedural terms. The semantics of these modals according to Klinge (1993) consists of two parts, both procedural. First, common to all of the modals is the notion of potentiality, by which Klinge (1993) means that all of the modals represent a situation or event as having the potential either to be or not to be an actual, verified situation or event (a WORLD SITUATION -distinct from a 'world event'). Klinge proposes a unitary account of the semantic category MODAL, as represented in figure 14, first as a generalisation, and second as a representation of (4.21):

- (4.21) a. John [modal] buy Mary's ticket.  
 b. MODAL [JOHN BUY MARY'S TICKET]

Figure 14: the semantics of MODAL



(Klinge 1993:324, 325)

In terms of the cognitive domains posited above, all the modals encode procedural information instructing an addressee to override the usual default assumption that situations and events described are realis. (actual, verified), and interpret any SR within the scope of a modal auxiliary as irrealis.<sup>54</sup>

The second element of the semantics of modals concerns the modals individually. Each individual modal is characterised as encoding procedural information which stipulates the potential relation between a SR within its scope and reality. For example, Klinge (1993:335-6) draws on a suggestion from Van der Auwera (1986)

<sup>54</sup> See §3.2.1 for a discussion of the differences between Klinge's definition of SRs and that proposed in this thesis.





Klinge (1993) undoubtedly provides an insightful and innovative analysis of the central modals, and the analysis of *shall* in §6.2 is heavily indebted to it. However, it is not without its problems. One relatively minor point is that it is difficult to see how the framework proposed for the general category MODAL could accommodate the other modals (*could, might, would, should* and *ought*) without proposing a single underlying lexical semantics for each of *will/would, shall/should* etc., which is in fact what Klinge implies (1993:348; p.c. 7/10/94). Were this approach adopted, a response would have to be formulated to Warner (1993), which argues that what looks like the verbal category 'finite' (e.g. *should = shall + past tense*) is not in fact inflectional in auxiliaries, but lexically specified, hence *should* is semantically independent of *shall*.

A more fundamental problem with Klinge's analysis, from the perspective of this thesis at least, is that a procedural characterisation of the semantics of the modals (I exclude *will* and *shall* for the time being) based on a general category MODAL characterised as 'potential', is untenable. The reason for this is that any account which incorporates the notion of 'potential' (or irrealis) into the semantics of *must* faces problems in accounting for the use of *must* in the expression of analytic propositions such as those repeated below:

(4.10)           A bachelor must be unmarried.

(4.11)           A bilabial nasal must be made with the lips closed.

Analytic propositions are not potential since their very expression constitutes their verification. The interpretation assigned to *must* in utterances of analytic sentences is akin to the logical modal operator L. The definition of L in terms of the other primitive modal logical operator M ( $LA = df \sim M \sim A$ ) is close to Klinge's paraphrase of the semantics of *must*: "The SITUATION REPRESENTATION does not turn out not to be a true description of a WORLD SITUATION" (1993:351), except that this whole paraphrase is subject to the additional feature 'potentiality', which is absent from analytic sentences. If potentiality plays no part in the semantics of *must*, then the sense of potentiality in the interpretation of sentences such as (4.12) must be inferred:

(4.12)           *Knock, knock*            "That must be the postman."

An utterance of (4.12) would normally be interpreted as a deduction from partial evidence (e.g. the postman always knocks twice, etc.), and as such is unverified in contrast to a conclusion based on direct acquaintance. (How the interpretation process functions in such cases will be discussed in §4.2.3 and §4.2.4 below.)

The use of *must* in the expression of analytic propositions often occurs in the protasis (the *if*-clause) of conditional sentences. Situations and events represented in the protasis are presented as hypothetical, and hence the default assumption concerning their existential status is that they are unverified (irrealis) in contrast to the usual default assumption that situations and events described are realis. For example, the speaker of (4.24) is not claiming that it is raining, but simply hypothesising about the consequences of going out if it were to rain:

(4.24)           If it rains, we will get wet.

In contrast, if an analytic proposition is presented in the protasis, it is necessarily verified. The use of *must* in the protasis helps the addressee to override the assumption set up by the *if*-clause that hypothetical events are irrealis. It does this by representing the proposition with which it is associated as being based on a deduction, thereby drawing the addressee's attention to its analytic character. In (4.25), for example, the assumption is that the sum of all three corners in triangles is 180°, rather than just that it might be:

(4.25)           If the sum of all three corners in triangles must be 180°,  
                  then there's no way I got the geometry test right.

[Papafragou 1995 (80a)]

The conclusion to be drawn is that Klinge's characterisation of the semantics of the general category MODAL as 'potential' plays no part in the semantics of *must*. As a result, if 'potential' does not figure in the semantics of *must*, then the existence of a unitary semantics for the general category MODAL is put in doubt. This does not entail that *must* is not procedural, simply that it cannot function procedurally by manipulating the interpretation of an SR in its scope relative to the irrealis domain. Arguments against a procedural analysis of the modals of any kind are given in Papafragou (1995), which is considered below.

#### 4.2.2 Modals as exponents of conceptual encoding: (Papafragou 1995)

Consideration of conditional sentences incorporating a modal auxiliary, such as (4.25), leads Papafragou (1995) to conclude that, when interpreted epistemically, modal auxiliaries encode conceptual information contributing to higher-level explicatures. Her argument is recreated here. In §2.3.2, higher-level explicatures were defined as "conceptual representations, capable of entailing and contradicting each other and representing determinate states of affairs [which] though true or false in their own right ... do not generally contribute to the truth conditions of their

associated utterances." (Wilson & Sperber 1993a:16) Since higher-level explicatures do not contribute to the truth-conditions of their associated utterances, placing epistemic *must* and *may* within the scope of logical operators results in infelicity; (4.25) expresses dynamic modality and (4.26) expresses epistemic modality:

(4.25) If the sum of all three corners in triangles must be  $180^\circ$ , then there's no way I got the geometry test right.

(4.26) ? If John must have a high IQ, then his teachers should treat him carefully.

[Papafragou 1995 (80a) and (81a)]

The use of *must* in (4.25) receives a dynamic interpretation - it is a statement about the way things are in the world without any expression of the speaker's commitment to the truth of the assertion; hence, there is only a single antecedent proposition in (4.25) which therefore has a well-formed logical structure of the form  $p \rightarrow q$ , where  $p$  = 'the sum of all three corners in triangles must be  $180^\circ$ ' and  $q$  = 'there's no way I got the geometry test right'. In (4.26) however, only an epistemic interpretation can be recovered, since 'John must have a high IQ' can be neither deontic (? 'John is obliged to have a high IQ') nor dynamic (? 'John is predisposed to have a high IQ'). (4.26) is infelicitous according to Papafragou (1995) because an epistemic interpretation does not contribute to the truth conditions of the associated utterance, that is to the truth conditions of  $p$  in the argument  $p \rightarrow q$ . This can be accounted for by viewing an epistemic interpretation as giving rise to a higher-level explicature; this higher-level explicature has its own truth conditions, but these are not part of  $p$  and therefore do not enter into the relation of entailment from  $p$  to  $q$ .

Similarly, treating the epistemic interpretation of *may* as contributing to higher-level explicatures could explain why (4.27a) below is felicitous if interpreted deontically (4.27b) but not epistemically (4.27c):

- (4.27) a. If I may go to the Ball, I will meet Prince Charming.  
 b. If I am permitted to go to the Ball, I will meet Prince Charming.  
 c. ? If it is possible that I go to the Ball, I will meet Prince Charming.

(4.27a) has the form  $p \rightarrow q$ . Under a deontic interpretation, where  $p^1$  = 'I am permitted to go to the Ball', *may* contributes to the truth conditions of  $p^1$  which can therefore enter into a relation of entailment with  $q$  ('I will meet Prince Charming'). With an epistemic interpretation, on the other hand, there are two sets of truth-conditions: one associated with  $p^2$ , ('I will go to the Ball'), and another associated with a higher-level explicature, ( $p^3$ ) '(I believe) it is possible that  $p^2$ '. Under an

epistemic interpretation, (4.27a) has the form  $p^3 \rightarrow q$ , rather than  $p^2 \rightarrow q$ , and since  $p^3$  does not entail  $q$ , an epistemic interpretation of (4.27a) is infelicitous.

On the basis of such considerations, Papafragou (1995) concludes that epistemic interpretations of utterances containing *must* and *may* are derived from conceptual information encoded by *must* and *may* contributing to higher-level explicatures; this same conceptual information in another utterance may be interpreted as contributing to the utterance expressed. She distinguishes between the epistemic and the non-epistemic uses of *must* and *may*, not in terms of any semantic difference between an epistemic and a non-epistemic *must* or *may* (this would be to fall back into polysemy), but in terms of differences in the logical scope of *must* and *may* when interpreted epistemically rather than non-epistemically. In the discussion of the illocutionary adverbials *seriously*, *frankly* and *confidentially* (2.3.2), it was shown that in indirect reported speech these are interpreted as their "synonymous manner-adverbial counterparts" (W&S 1993a:17) which contribute to the proposition expressed rather than to higher-level explicatures. It is not the semantics of *seriously*, *frankly* or *confidentially* which changes, but their scope relations relative to other elements of the associated utterance (encoded or inferred). Similarly, Papafragou (1995) proposes that it is not the semantics of *must* or *may* which changes when a deontic or dynamic interpretation is recovered rather than an epistemic one, but simply the scope relations between *must* or *may* and other elements of the utterance. On an epistemic interpretation, *must* or *may* is outside the proposition ( $p$ ) and therefore cannot enter into an entailment relation of the form  $p \rightarrow q$ . On a non-epistemic interpretation, however, *must* or *may* is inside the proposition  $p$  and can thus fall within the scope of the logical operator  $\rightarrow$ . This suggests that epistemic and non-epistemic interpretations are derived from distinct logical forms.

There are problems with Papafragou's (1995) arguments, however. First, it is not always the case that the use of *must* or *may* with an epistemic interpretation in the protasis of a conditional is infelicitous:

(4.28)        If Dhara may be here for another week, I won't invite Alison to stay.

The situation is one in which Dhara, our lodger has informed us that it will be at least a week before his houseboat is in York; this leads me to believe that he will be with us for at least another week and so the spare room will be occupied. The epistemic interpretation of *may* in (4.28) clearly contributes to the proposition which entails 'I won't invite Alison to stay'. Since it is not correct to say that in this case the epistemic interpretation of *may* does not contribute to the proposition expressed, it is doubtful

whether this epistemic interpretation is a higher-level explicature at all, given Wilson & Sperber's definition (1993a:16).

Returning to the earlier examples, the reason an epistemic reading of (4.27a) is infelicitous is simply that  $q$  ('I will meet Prince Charming') does not follow from  $p$  ('it is possible that I go to the Ball'). If 'might' is substituted for 'will' in  $q$ , an epistemic reading becomes acceptable:

- (4.27) a. ? If I may go to the Ball, I will meet Prince Charming.  
c. ? If it is possible that I go to the Ball, I will meet Prince Charming.  
(4.29) a. If I may go to the Ball, I might meet Prince Charming.  
b. If it is possible that I go to the Ball, I might meet Prince Charming.

Where *must* occurs in the protasis of a conditional sentence, its unacceptability is not due simply to the fact that it is epistemic; compare (4.26) with epistemic *must*, and (4.30) with epistemic *may*:

- (4.26) ? If John must have a high IQ, then his teachers should treat him carefully.  
(4.30) If John may have a high IQ, then his teachers should treat him carefully.

So, rather than claim, as Papafragou (1995) does, that an utterance with *must* or *may* encodes higher-level explicatures when epistemically interpreted (which don't enter into entailment relations when placed within the scope of a logical operator) whilst non-epistemic *must* and *may* encode propositional information (capable of entering into entailment relations), I propose that the identification of a particular type of modality (epistemic, deontic or dynamic) is inferred. Specifically, I propose that modals are operators with scope over SRs. Once a SR has been interpreted in accordance with the procedural information encoded by an operator, various explicatures and implicatures may be recovered; these may concern higher-level propositional attitudes, for example the degree of speaker commitment to the resulting proposition or speech act information, resulting in epistemic or deontic interpretations.

In order to predict the nature of the implicatures inferred from utterances containing a given modal, an explicit semantics of that modal is required which licenses the relevant interpretation. Such an analysis is provided by Groefsema (1995a), at least for the modals *can*, *may*, *must* and *should*. I shall briefly review Groefsema's account, with the following aims in mind: first, that it provides the basis of a counter-argument

to one final claim in Papafragou (1995) in support of the view that epistemic interpretations of modals are higher-level explicatures; secondly, that it provides a viable alternative to the 'possible worlds' model of modality; and finally, that it can be recast in explicitly procedural terms (although without reference to cognitive domains).

#### 4.2.3 Towards a procedural analysis of the modals: (Groefsema 1995a)

Groefsema (1995a:62) characterises the basic meanings of four of the modals as follows:

*Can:*  $p$  is compatible with the set of all propositions which have a bearing on  $p$ .

*May:* There is at least some set of propositions such that  $p$  is compatible with it.

*Must:*  $p$  is entailed by the set of all propositions which have a bearing on  $p$ .

*Should:* There is at least some set of propositions such that  $p$  is entailed by it.

(where  $p$  is the proposition expressed by the rest of the utterance).

Informally, the propositions which have a bearing on  $p$  constitute the evidence for  $p$ . This could be further refined as stating that they constitute the evidence for the set of assumptions according to which the optimal relevance of  $p$  is determined. *Must* expresses that all the evidence entails  $p$ , whilst *should* expresses that at least some of the evidence entails  $p$ ; *can* expresses that all the evidence does not entail  $\sim p$ , whilst *may* expresses that at least some of the evidence does not entail  $\sim p$ .

Groefsema (1995a) explains in some detail the role played by the above characterisations of the modals in the utterance interpretation process, and this will not be repeated here, except in regard to the relation between negation and epistemic interpretations of *must* and *may*. Papafragou (1995:4) argues that epistemic uses of *must* and *may* do not exhibit compositionality with elements belonging to the proposition expressed, such as negation (*need not* and *cannot* may be used instead), and that this provides evidence for viewing epistemic *must* and *may* as being outside the proposition (and hence as contributing to higher-level explicatures):

(4.31) a. \* He must not find Mozart exciting. (epistemic)

b. He need not find Mozart exciting.

(4.32) a. \* He may not be his tutor. (epistemic)

b. He cannot be his tutor.

[Papafragou 1995 (84) and (83)]

Groefsema (1995a:74-8) accounts for this phenomenon as follows, first in relation to *must* and then in relation to *may*. When an utterance containing *must* is negated, the negation affects the main predication, as in (4.33), rather than the modal predication as in (4.34):

(4.33)  $\sim p$  is entailed by the set of all propositions which have a bearing on  $p$ .

(4.34) It is not the case that  $p$  is entailed by the set of all propositions which have a bearing on  $p$ .

[Groefsema 1995a (49) and (48)]

This explains why the negation of (4.35a) is not (4.35b):

- (4.35) a. John must be thirty.  
b. ? John mustn't be thirty.

(4.35b), on an epistemic interpretation with the logical form of the kind given in (4.34) where  $p$  = 'John be thirty', is infelicitous simply because it is not very informative. It leaves open the possibility that some but not all of the evidence entails  $p$ , or that  $p$  is compatible with some of the evidence, or that  $p$  is incompatible with some of the evidence. These propositions could be expressed by using *should*, *may* and *can't* respectively:

(4.36) John should be thirty.

(4.37) John may be thirty.

(4.38) John can't be thirty.

(4.35b) is not used because it is subsumed under any of (4.36) to (4.38), and therefore a speaker aiming for optimal relevance will utter the more informative (4.36), (4.37) or (4.38) rather than (4.35b). When an utterance containing *must* is negated, only a non-epistemic interpretation with the logical form given in (4.33) is possible. Why then can't the negative of epistemic *must* be expressed using *must not* with the logical form in (4.33)? Groefsema (1995a:76) suggests that if the speaker knows that  $\sim p$ , then there is no need to draw the addressee's attention to any possible evidence for  $p$ , and relevance is best served simply by stating  $\sim p$ ; if, on the other hand, the speaker knows that  $\sim p$  because there is some piece of evidence which is incompatible with  $p$ , this is more precisely expressed through the use of *can't*, hence *can't* supplies the missing negative for epistemic *must*.

Where *may not* is concerned, epistemically it negates the main verb, as in (4.39), while deontically the modal is negated, as in (4.40):

(4.39) There is at least some set of propositions such that  $\sim p$  is compatible with it.

(4.40) It is not the case that there is at least some set of propositions such that  $p$  is compatible with it.

[Groefsema 1995a (58) and (59)]

Groefsema accounts for this as follows:

"Imagine that Mary has been invited to a party, and she thinks that she has to go, but doesn't feel like it. Her mother walks in and says:

(4.41) You may not go to the party.

Because a logical form along the lines of (4.40) does not yield a relevant interpretation (i.e. Mary has the assumption that she has to go), Mary is entitled to go for the logical form (4.41<sup>˘</sup>):

(4.41<sup>˘</sup>) There is at least some set of propositions such that [ $p$  it is not the case that I go to the party] is compatible with it.

This then communicates to Mary that her mother has reason to think that it may be alright for her not to go (i.e. a possibility interpretation).

Now imagine a situation in which Mary has been invited to a party, and she wants to go, but she expects that her parents won't let her. Again her mother walks in and says (4.41). In this case the logical form in (4.41<sup>˘</sup>) will not yield a relevant interpretation because Mary already knows that there is some evidence which is compatible with her not going to the party so that the only interpretation she can go for is based on the logical form in (4.41<sup>˘˘</sup>):

(4.41<sup>˘˘</sup>) It is not the case that there is at least some set of propositions such that [ $p$  I go to the party] is compatible with it.

[...] Because Mary wants to go to the party, the logical form in (4.41<sup>˘˘</sup>) gives rise to a contradiction. However, because it is her mother speaking to her, assumptions about her authority become easily available so that Mary can conclude ... that her mother is not allowing her to go to the party [and] that her own wishes are not taken into consideration."

(Groefsema 1995a:77-8)

In §4.2.4, I shall bring Groefsema's account of *can*, *may*, *must* and *should* into line with the model of procedural encoding proposed in §3.2; that is, I shall demonstrate how these modals can be characterised as operators with scope over SRs whilst retaining the insights into their behaviour captured by Groefsema (1995a). This reformulation will be more explicit in that it captures the distinction between phases two and three of utterance interpretation. I shall not attempt, however, to extend the



analysis provided in Groefsema (1995a) to other modals; Groefsema (1995a) characterises *can*, *may*, *must* and *should* in terms of a closed set of oppositions which defines the semantics of these modals as expressing that *p* is compatible with/entailed by all/some of the propositions that have a bearing on *p*. It is not clear how this system could be extended to accommodate the other modal auxiliaries.

#### 4.2.4 A procedural analysis of *can*, *may*, *must* and *should*

In order to bring Groefsema's account into line with the procedural model described above, I shall begin by reading SR where Groefsema (1995a) has *p*. In my analysis, what Groefsema (1995a) terms *p* is not fully propositional since its existential status and (often) its temporal reference are not known prior to enrichment in accordance with the information encoded by a modal; Groefsema (1995a), on the other hand, treats *p* as fully propositional. For example, (4.42) has the logical form in (4.43) which is explicitly given as (4.43') according to my analysis, but according to the analysis provided in Groefsema (1995a:66) the logical form of (4.42) is (4.43''):

(4.42) John may smoke.

(4.43) MAY [JOHN SMOKE]

(4.43') There is at least some set of propositions such that [SR John smoke] is compatible with it.

(4.43'') There is at least some set of propositions such that [*p* John smoke at time *t*] is compatible with it.

Now, [JOHN SMOKE] is not a proposition, since it is indeterminate between a realis, atemporal reading [John smokes habitually], an irrealis, atemporal reading [potentially John smokes habitually] and a future time (and therefore irrealis) reading [potentially John smokes at  $t_1$  where  $t_1 > t_0$ ]. Logical forms derived from utterances lacking explicit temporal adverbials do not give rise to propositions such as [John smoke at time *t*] until phase two of utterance interpretation has occurred. Thus the modals constrain the interpretation of SRs rather than propositions and will therefore be treated as operators according to the definition given in 3.2.3, repeated below:

An operator is any expression which encodes procedural information concerning the interpretation of a SR.

As with all procedural encoding, the modals constrain the inferential phases of utterance interpretation. Inference in phase two of utterance interpretation involves interpreting SRs and elements within SRs (such as the conceptual information encoded by nominal referring expressions) with respect to the utterance interpretation context.

The mandatory inferential processes at phase two of utterance interpretation include reference assignment (of nominal referring expressions to discourse entities) and the interpretation of SRs relative to cognitive domains (basic or abstract). The modals (*can*, *may*, *must* and *should*) do not constrain the interpretation of SRs with respect to any domains (at least not directly), but rather with respect to the other assumptions which constitute the utterance interpretation context (the propositions which have a bearing on the SR, or the 'evidence') and this occurs at phase two of utterance interpretation. At this phase, no particular assumptions are identified by the addressee; the semantics of the modals are underdetermined in this respect, and merely state that some or all of the assumptions in the utterance interpretation context are compatible with or entail the SR. However, at phase three, an addressee may identify a particular assumption and thereby enrich the proposition to which the modal contributes to recover the intended contextual effects, including, where appropriate, the type of modality expressed. This process can be illustrated with regard to the interpretation of (4.14), repeated below, which gives rise to the logical form in (4.44) with the explicit paraphrase (4.44<sup>^</sup>):

(4.14) Freddie can recite a poem.

(4.44) CAN [FREDDIE RECITE A POEM]

(4.44<sup>^</sup>) [SR Freddie recite a poem] is compatible with the set of assumptions which have a bearing on [SR Freddie recite a poem].

The precise interpretation of (4.14) depends on the utterance interpretation context and the assumptions contained in it. If the utterance interpretation context includes the assumptions that the speaker is Freddie's father and that Freddie is asleep in bed, (4.14) will be interpreted as (4.19) in which the modal *can* is inferentially enriched to yield a dynamic interpretation:

(4.19) Freddie has the ability to recite a poem.

If the utterance interpretation context includes the assumptions that the speaker is Freddie's teacher, Freddie has the ability to recite a poem and Freddie wants to recite a poem in the school play, the crucial assumption is that the speaker is giving permission for Freddie to recite his poem, and (4.14) will be interpreted as (4.20) in which *can* is inferentially enriched to yield a deontic interpretation:

(4.20) Freddie is permitted to recite a poem.

In this reformulation of Groefsema's analysis, the semantics of *can*, *may*, *must* and *should* are not connected directly with the irrealis domain, but with assumptions in

the utterance interpretation context. My claim is that when a situation or event expressed through an utterance containing one of these modals is interpreted as irrealis, this is inferred. The modals *can*, *may*, *must* and *should* do not encode procedural information constraining this inference, that is, the relation of the SR to the irrealis domain is not procedurally encoded by these modals. We have already seen that *must* is not always used in the expression of irrealis situations, since it can be used to express analytic propositions. Even *may* is used to talk about situations and events which are not irrealis:

(4.9) He may be a university professor, but he sure is dumb.

If (4.9) were uttered in a context where both the speaker and the addressee know that 'he' refers to Gordon and that Gordon is a university professor, *may* is used in the expression of an actual, verified and therefore realis proposition [Gordon is a university professor] without infelicity. How is this achieved? The clause 'He may be a university professor' gives rise to the logical form in (4.45):

(4.45) There is at least some set of assumptions such that [<sub>SR</sub> Gordon be a university professor] is compatible with it.

As it stands this is highly uninformative, since it is mutually manifest that Gordon is a university professor. It is the addition of the following clause 'but he sure is dumb', that indicates how (4.9) is intended to achieve relevance. The conjunct *but* signifies that the following proposition is intended to weaken any conclusions drawn from the preceding proposition (in 1.4.1 we saw that new information can achieve contextual effects through the strengthening of existing assumptions; *but* indicates that new information to be presented is intended to weaken rather than strengthen existing assumptions). From an utterance of 'he sure is dumb' the addressee can easily deduce [Gordon is not intelligent]. Now, [Gordon is intelligent] is one of the assumptions compatible with [Gordon is a university professor], so the addressee can infer that the speaker wishes to convey that one particular assumption, namely [Gordon is intelligent], is not compatible with the assumption [Gordon is a university professor]. Thus, (4.9) is interpreted as follows:

(4.46) There is at least some set of assumptions such that [Gordon is a university professor] is compatible with it, but the assumption [Gordon is intelligent] is not a member of this set.

This said, utterances of analytic sentences and sentences like (4.9) are rare in comparison with utterances in which modal expressions are used to represent irrealis

situations and events; SRs within the scope of a modal are usually interpreted as representations of irrealis situations and events. This is the result of inference (functioning on the basis of the principle of relevance and the utterance interpretation context) not of any procedural information encoded by *can*, *may*, *must* or *should*.

### 4.3 Summary and Prospects

#### 4.3.1 Summary

I began by demonstrating that directly truth-conditional accounts of the modal auxiliaries, whether logical or descriptive, fail to account for their interpretation and distribution. The problem with such accounts is that they do not recognise that modal auxiliaries significantly underdetermine the interpretation of utterances in which they occur, and that utterance interpretation is a largely inferential process as regards modality, as it also is with temporal reference and aspectual information. I therefore proposed that the function of the modal auxiliaries might parallel that of tense and aspect markers, by constraining the inferential determination of modality, just as tense and aspect markers constrain the interpretation of temporal reference and aspectuality. Since the specification of the modal status of a situation or event contributes to the proposition expressed, it is the manipulation of sub-propositional conceptual representations of situations and events which the modal auxiliaries constrain. Like tense and aspect markers therefore, modal auxiliaries could also be characterised as operators, that is, exponents of procedural encoding at phase two with scope over situation representations.

If the modals constrain the manipulation of SRs, the parameters in relation to which this occurs must be stipulated. By discussing and rejecting Klinge's (1993) proposal that the modals *can*, *may* and *must* encode potentiality, I demonstrated that if these modals do in fact constrain the manipulation of SRs, it is not with respect to the irrealis cognitive domain. In contrast, Papafragou (1995) argued that, when interpreted epistemically, *may* and *must* encode conceptual information contributing to higher-level explicatures. Papafragou's analysis (which is based on a possible worlds semantics, against which I argued in §4.1.1) was rejected when it was shown that epistemic modality need not fall outside of propositional scope. Instead, I suggested that *can*, *may* and *must* encode constraints on the interpretation of any SR within their scope, and that once a SR has been interpreted in accordance with these constraints, other information, such as whether the modality is epistemic, deontic or dynamic, may be inferred.

The precise nature of the constraints on interpretation encoded by *can*, *may*, *must* and also *should* were determined by a reanalysis in procedural terms of Groefsema (1995a):

- Can:*  $p$  is compatible with the set of all propositions which have a bearing on  $p$ .  
*May:* There is at least some set of propositions such that  $p$  is compatible with it.  
*Must:*  $p$  is entailed by the set of all propositions which have a bearing on  $p$ .  
*Should:* There is at least some set of propositions such that  $p$  is entailed by it.

where  $p$  was redefined as the SR over which the modal operator has scope, and the propositions which have a bearing on  $p$  were specified as those which constitute the evidence for the set of assumptions according to which the optimal relevance of  $p$  is determined.

This analysis accounted for the data which falsified both Klinge (1993) and Papafragou (1995). Unlike Papafragou (1995) the analysis presented in §4.2.4 does not depend on reference to a possible worlds model, and characterises *can*, *may*, *must* and *should* in purely procedural terms. Unlike Klinge (1993) reference to potentiality or the irrealis cognitive domain does not constitute part of the semantic ontology of any of the four modal auxiliaries discussed; rather, *can*, *may*, *must* and *should* relate SRs in their scope to assumptions in the utterance interpretation context, leaving irrealis to be inferred where appropriate.

#### 4.3.2 Prospects

In this chapter, I have proposed an analysis of *can*, *may*, *must* and *should* as exponents of procedural encoding at phase two of utterance interpretation with scope over situation representations. The account offered here, like the other accounts discussed, characterises each modal auxiliary as a member of a set, and recognises that the interpretation and distribution of each member of that set can not be divorced from the dynamics of the set as a whole. There are, however, certain properties of *should* and *must* in particular, which are not accounted for by Groefsema's analysis or by my reformulation of it (nor by any other analysis which I have encountered).

*Should* is distinguished not only from *can*, *may* and *must*, but from all of the modals by its 'optimistic' nature, which renders (4.47a) but not (4.48a) felicitous, in contrast to the acceptable (4.47b) and (4.48b) where other modals are used:

- (4.47) a. The Linguistics Department fancy dress party should be fun.  
 b. The Linguistics Department fancy dress party may/could/will be fun.
- (4.48) a. ? The Linguistics Department fancy dress party should be boring.  
 b. The Linguistics Department fancy dress party may/could/will be boring.

This 'optimistic' interpretation of *should* is overridden in the protasis of conditional sentences, where other modals are usually infelicitous:

- (4.49) a. If it should rain, we will have to stay at home.  
 b. ? If it may/could/will rain, we will have to stay at home.

Certain features of the semantics of *must* manifest themselves most clearly as differences in interpretation between *must* and *have to*. Apart from the differences concerning the relative scope of these expressions and negation, which may be syntactic in origin, the interpretation of utterances containing *must* and *have to* differ in the following ways:

- (4.50) a. You must be very intelligent to understand Labelled Deduction.  
 b. You have to be very intelligent to understand Labelled Deduction.
- (4.51) a. I must find out more about Labelled Deduction.  
 b. I have to find out more about Labelled Deduction.

(4.50a) suggests that the addressee does understand Labelled Deduction (and is therefore very intelligent), whereas (4.50b) conveys no such suggestion, the interpretation being simply that to understand Labelled Deduction it is necessary to be very intelligent. Other differences in interpretation arise between (4.51a) and (4.51b); whereas (4.51a) suggests that the compulsion to find out more about Labelled Deduction is internal, for example the speaker's own curiosity, (4.51b) suggests an external compulsion, say, an examination syllabus.

Also not discussed, are variations in usage concerning the modal auxiliaries and their historical development. (Such considerations fall beyond the scope of the preceding analysis, which is concerned purely with the procedural information encoded by the modals *can*, *may*, *must* and *should*.) Linguistic variation and change are closely linked, and will be discussed in chapter 5 in relation to the conceptual/procedural distinction at phase two. In chapter 6 an account of the grammatical markers *will*, *shall* and *be going to* will be provided which incorporates both synchronic and diachronic considerations.

## 5 GRAMMATICIZATION AND PROCEDURAL ENCODING

Grammaticization (also known as 'grammaticalization'), a term introduced by Meillet (1912), refers to the process whereby "words from major lexical categories, such as nouns, verbs and adjectives, become minor, grammatical categories such as prepositions, adverbs and auxiliaries, which in turn may be further grammaticalised into affixes." (McMahon 1994:160)

Grammaticization results in semantic, syntactic, morphological and phonological changes in the grammaticizing expression ('gram'). Semantically, the developing gram moves from encoding a specific semantic content, appropriate to a restricted range of contexts, to encoding a very generalised, reduced semantic content, appropriate in an increasingly wide range of contexts, and increasingly dependent on contextual factors for its interpretation. This is termed 'semantic bleaching' (following Givón 1975a). An example of semantic bleaching is afforded by the change from Old English *willan* with a specific sense of desire and volition and selecting only animate subjects, to Modern English *will* with a more general meaning and highly variable interpretation with no selection restrictions on appropriate subjects. The increased dependence of a gram on context for its interpretation can eventually lead to redundancy, as in the co-occurrence of the past tense suffix *-ed* with temporal adverbials expressing past time reference.

Syntactically, increased grammaticization leads to increased rigidity in syntactic position and scope relations (see §3.2.1), leading in some cases to affixation. Affixation is frequently accompanied by phonetic reduction as the gram ceases to be stressed, as in the bound allomorphs of *will*: as a suffix to a subject NP [ɪ] (e.g. *you'll*) and affixed to the reduced allomorph of *not* (itself a gram), as *won't*. Affixation is not a prerequisite for phonetic reduction as *gonna* [gɒnə] - the reduced form of *going to* - illustrates (the accompanying form of the auxiliary *be* in the *be going to* construction also tends to be affixed to the preceding NP with accompanying phonetic reduction).

The specific type of grammaticization to be discussed here is the process whereby operators (grams encoding procedural information constraining the inferential interpretation of SRs within their scope) develop from lexical verbs. Operators, such as the tense, aspect and modality markers discussed in chapter 3 are all grams, and so, from the perspective of the conceptual/procedural distinction, grammaticization involves a shift from conceptual encoding to procedural encoding in a single expression over time. Given the arguments concerning inferential enrichment of elements within logical forms (§1.3.2) and the problems encountered by polysemous

accounts of the modal auxiliaries (§3.4.2), I shall maintain a unitary, monosemous semantics for every operator.

The model of procedural encoding developed in this thesis must be reconciled with the following aspects of grammaticization. First, a clear distinction has been drawn between conceptual and procedural encoding; there is no information type intermediate between conceptual and procedural information. This claim must be reconciled with the fact that formal grammaticization is a gradual process with many individual expressions occupying intermediate positions on a continuum between fully lexical and fully grammaticized, whilst there exist modal expressions in languages other than English, which behave semantically in a similar way to the English modal auxiliaries (that is they often give rise to similar interpretations, lack selection restrictions on permissible subjects, are highly context dependent, etc.) yet which do not exhibit the formal features of grammaticization. Secondly, procedural encoding has been characterised as constraining the inferential phases of utterance interpretation. We would therefore expect to see a link between the onset of grammaticization and the inferential processes of phase two of utterance interpretation; I shall therefore investigate the causes of grammaticization. Finally, I will take into consideration the fact of semantic retention, whereby conceptual information encoded by the lexical source expression of a gram is still accessible in the use of that gram in certain contexts (e.g. the volitional interpretation of *will* which may be recovered in questions: *Will you do the washing up?* i.e. 'Are you willing to do the washing up?'). I shall address each consideration in turn.

I shall exemplify grammaticization with reference to the modal auxiliary *will*. In §3.2.1, *will* was cited, without any attempt at justification, as an example of an operator; it was simply assumed, on the basis of surface syntactic structure, that *will* encodes procedural information with wider scope than tense and aspect markers. The motivation for viewing *will* as an operator (encoding procedural information constraining the manipulation of sub-propositional conceptual representations) derives from the evidence presented in §4.1 that the modals are not directly truth-conditionally interpretable. I shall continue to discuss *will* in chapter 6.

## **5.1 The gradualness of grammaticization**

### **5.1.0 Introduction**

It is generally accepted within the literature that grammaticization is a gradual process (see for example Lichtenberk 1991:37, Hopper & Traugott 1993:94-129, and Bybee



*et al* 1994:6). The shift from lexical to grammatical expression is often viewed as a continuum, along which grammaticized and grammaticizing expressions are distributed, clustering at certain points (as auxiliaries or affixes for example). It is further assumed that the various changes involved in grammaticization - both formal and semantic - occur in parallel:

"both types of formal change in grammaticization parallel the main types of semantic change in grammaticization. Phonetic reduction - the loss of specific phonetic properties - parallels [semantic] reduction or generalization, which is also the loss of specific properties. The fusion of a developing gram to adjacent lexical material in affixation is parallel to the growing functional dependence of grams and their conceptual cohesion with lexical stems."

(Bybee *et al* 1994:106)

In this section I shall suggest that this characterisation of grammaticization as both formally and semantically gradual has unfortunate consequences for the possibility of systemic, synchronic accounts of grams and for the conceptual/procedural distinction. In response to this problem I shall present evidence that grammaticization, whilst clearly gradual as far as formal change is concerned, is not a semantically gradual process.

#### 5.1.1 The consequences of semantically gradual grammaticization

In chapter 4, I proposed an analysis of the modal auxiliaries based on Groefsema (1995a) which was synchronic and structuralist in that it accounted for the semantics of individual modal auxiliaries in terms of their membership of a system of oppositions (although in §4.3.2 I noted that certain properties of *should* and *must* were not explained by this analysis). If, as has been claimed, grammaticization is a gradual process, both formally and semantically, the possibility of such an analysis is put in doubt for the following reasons: If each modal auxiliary is undergoing a process of gradual grammaticization, it will be inherently semantically unstable; any system of oppositions between the modal auxiliaries (such as that proposed in Groefsema 1995a) must therefore also be unstable, and hence resistant to systematic, synchronic analysis.

It is certainly true that, if not the semantics, at least the distribution of the modal auxiliaries has been subject to change, as the following uses of *must* from 1885 illustrate:

- (5.1) For four or five days his condition was most crucial; indeed, I believe firmly that had it not been for Foulata's indefatigable nursing he must have died ...  
For two days we thought that he must die, and crept about with heavy hearts.

(H. Rider Haggard 1885 *King Solomon's Mines* London, etc., Cassell p.217)

In present-day English *he would have died* and *he would/was going to die* respectively are more appropriate, but the interpretation of the 1885 usage poses no problems.

Bybee (1988:247) rejects the structuralist view that grams are assigned a value by the oppositions they enter into, on the basis that grams develop gradually out of lexical material and therefore have inherent semantic content. She argues that since there are consistent cross-linguistic and diachronic relationships between lexical and grammatical meaning, the diachronic source of a gram must be more important in determining its meaning than the existence of other grams in the language. This is reflected in the observation that similar lexical sources for grams are found across languages, but systems of grams differ cross-linguistically. Bybee (1988:253) notes that a developing gram must constrict the distribution of other grams since "every time it is used another gram is not." However, she claims that when this happens a developing gram does not imbue another gram with meaning, and illustrates her argument with reference to the effect of the development of *will* on *shall* (*shall* being the earlier of the two forms). The specifics of this argument will be addressed in §6.2 when *will* and *shall* are discussed in detail, but a few general responses to Bybee (1988) can be offered now.

Whilst the lexical source of a gram undoubtedly has an important influence on the semantics of the resulting gram, the system of oppositions into which the gram enters also has a significant effect on the semantics and the distribution of the gram. Take the case of cognates in different languages which develop divergent meanings, such as Modern English *will* and Modern German *wollen*. *Will* is rarely the best translation for a form of *wollen*, rather, *want*, *like* or some other verb of desire or volition is more often appropriate. If English had not developed *want* (or an alternative form with a similar meaning) the development of *will* might well have paralleled that of *wollen*; but *will* and *wollen* are clearly semantically distinct, despite sharing a common lexical source. Bybee (1988) is therefore incorrect to claim that the diachronic source of a gram is more important in determining its meaning than systemic considerations; both are important. Furthermore, some distributional features of the modal auxiliaries, such as the interaction of epistemic *must* and *may* with negation (see §4.2.3), can only be explained by reference to a system. These considerations do not, however, answer the argument that a gradually developing semantics of a given gram can not be analysed in terms of a systemic, synchronic characterisation, such as that given for *can*, *may*, *must* and *should* in chapter 4.

In addition, the characterisation of grammaticization as a semantically gradual process is incompatible with the notion of a strict distinction between conceptual and procedural information types; if there is an intermediate position between lexical and grammatical encoding, and if these correspond to conceptual and procedural encoding respectively, as I have claimed, there must be an intermediate position between conceptual and procedural encoding. Rather than respond to this charge directly, I shall examine evidence which contradicts the underlying assumption that grammaticization is semantically gradual.

### 5.1.2 Counter-evidence from verb-serialisation

The underlying assumption that semantic development occurs in parallel with the formal aspects of grammaticization (Bybee *et al* 1994:106) is challenged in Givón (1991). Givón looked at verb-serialisation, whereby:

"An event/state that one language codes as a simple clause with a single verb is coded in another language as a complex clause with two or more verbs."

(Givón 1991:81)

Verb-serialisation can function not only as co-lexicalisation (for example, the translation of English *fetch* as *aller chercher* in French) but grammatically, for example by assigning nominal case-roles:

- |       |    |  |                |                 |
|-------|----|--|----------------|-----------------|
| (5.2) | a. | she <i>take-stick break</i><br>'she broke the stick'                       | (patient)      |                 |
|       | b. | she <i>walk go-market</i><br>'she walked <i>to</i> the market'             | (locative)     |                 |
|       | c. | he <i>work give-her</i><br>'he worked <i>for</i> her'                      | (benefactive)  |                 |
|       | d. | she <i>take-knife cut</i> meat<br>'she cut the meat <i>with</i> the knife' | (instrumental) | (Givón 1991:82) |

and by marking tense and aspect:

- |       |    |   |              |                 |
|-------|----|---|--------------|-----------------|
| (5.3) | a. | he <i>stay</i> work<br>'he <i>is</i> working'           | (durative)   |                 |
|       | b. | he <i>go</i> work<br>'he <i>will</i> work'              | (future)     |                 |
|       | c. | he work <i>finish</i><br>'he <i>has already</i> worked' | (perfective) | (Givón 1991:83) |

Simply by looking at the grammar of a serial verb language, it is impossible to determine whether speakers of that language perceive the events they describe using

serial-verbs as single events or 'multi-events' (i.e. a combination of distinct events). On the assumption that there is an iconic relation between grammar and cognition it has been argued (e.g. Pawley 1987) that multi-verb sequences therefore code multi-propositional sequences and hence refer to multi-events. However, it has been noted (e.g. Bradshaw 1982, Crowley 1987) that the same iconicity principle can be used to argue that serial-verbs represent single events, by invoking grammatical criteria which treat a cluster of serial verbs as a single constituent coding a single proposition. In both arguments, "one winds up with an inevitable circularity: Grammar is used first to *define* cognition, and then is said to correlate with it." (Givón 1991:86)

However, there is one major, well-attested iconicity principle which is dependent on neither grammar nor intuitive free translation (Givón 1991:119). That is the principle that the temporal-physical distance between stretches of linguistically encoded material, determined by rhythm, intonation contours and pauses in speech, correlates directly with cognitive 'packaging'; for example, the idea that pause length between linguistic expressions such as serial-verbs correlates with the conceptual 'distance' between the information encoded by those expressions. Givón (1991:86) reasoned that if serial-verb constructions reflect a multi-event cognitive segmentation of reality, then the pauses that characteristically occur at the boundaries of main clauses in non-serial-verb languages should also occur in serial-verb languages at serial-verb clause boundaries. Conversely, if serial-verbs represent single events, the probability of a pause occurring at a serial-verb clause boundary should be significantly lower than at a main clause boundary. Givón's findings are summarised below:

- (i) The probability of serial-verb constructions showing a pause is much lower, by a clear order of magnitude, [than] the pause probability associated with typical main clauses; and
- (ii) The pause probability of serial-verb constructions falls within the probability range of mid-clause pauses associated with lexical words, or is even lower, i.e. falling within the range of the probability of mid-word pauses between grammatical morphemes.

(Givón 1991:116)

What these findings demonstrate is that lexical expressions can be used with the same functional load as grammaticized alternatives, as indicated by the iconic features of the utterances in which they occur. This leads Givón to conclude that,

*"cognitively, grammaticalization is not a gradual process, but rather an instantaneous one.... The minute a lexical item is used in a frame that intends it as a grammatical marker, it is thereby grammaticalized."*

(Givón 1991:123)

When a lexical expression is used, as Givón puts it, in a frame that intends it as a grammatical marker, is the lexical meaning of the expression still recovered, and if

not, why not? The lexical semantics of the expression (that is, the conceptual information encoded) is clearly 'accessed' in addition to the newly encoded grammatical semantics (that is, the procedural information); since Swinney (1979) it has been accepted that alternative meanings of ambiguous words are 'accessed' or 'activated' even when a context selects for only one. The inappropriate meaning(s) of an ambiguous word are soon deactivated however, leaving only the appropriate lexical entry (sense) to be 'recognised' or 'recovered'; that is, in terms of the three phase model of utterance interpretation, only the appropriate meaning is fed into phase two to be inferentially enriched. In the case of a formally lexical expression used functionally/semantically as a gram, the newly encoded procedural information is automatically recovered since it provides a necessary processing constraint on the interpretation of the associated clause. If the resulting interpretation achieves adequate contextual effects on its own, the interpretation process should cease according to the principle of relevance, since recovering and inferentially enriching the conceptual information also encoded would increase processing effort and reduce relevance.

Givón (*ibid.*) goes on to suggest that the supposedly gradual nature of grammaticization is in fact the result of the gradual nature of the formal structural adjustments which follow (sometimes long afterwards) original, instantaneous developments at the functional level. The frequent delay in formal grammaticization, and its gradual nature when it does occur, explains the existence of expressions which exhibit the semantic characteristics of grams but the formal characteristics of lexical items (for example verbs in languages other than English which fulfil a similar semantic function to that of the English modal auxiliaries whilst behaving lexically as main verbs).

### 5.1.3 Counter-evidence from Kiswahili

A clear example of the functionally instantaneous but formally gradual nature of grammaticization is provided by the development of the Kiswahili aspectual marker *-mesha*. Tense, aspect and modality markers in Kiswahili are encoded as preverbal prefixes:

(5.4)            Ni + me + sóma        *I have read*  
                   SP PERF read

Kiswahili MTA prefixes are derived from auxiliary forms which in turn are derived from main verbs. This process is evidenced in contemporary Kiswahili in the recent development (mentioned in neither Ashton 1947 nor Polomé 1967) of the MTA

prefix *-mesha-* with the sense of *have already*, derived from the verb *kwisha* ('finish') plus the perfect marker *-me-*, itself derived from the (archaic) verb *mele* which also meant 'finish' (Nurse 1989:287). The examples below illustrate in turn the use of *me* + *kwisha* in a serial verb construction either with or without functional grammaticization (5.5), as a fully grammaticized prefix with the form *-mekwisha-* (5.6), and as a reduced prefix with the form *-mesha-* (5.7); each of these constructions is current in contemporary Kiswahili:

- |       |  |                       |  |
|-------|--|-----------------------|--|
| (5.5) | A + me + kwisha<br>SP PERF <i>kwisha</i>       | ku + soma<br>INF read | <i>She has finished reading</i><br><i>She has already read</i> |
| (5.6) | A + mekwisha + soma<br>SP <i>mekwisha</i> read |                       | <i>She has already read</i>                                    |
| (5.7) | A + mesha + soma<br>SP <i>mesha</i> read       |                       | <i>She has already read</i>                                    |

The construction in (5.6) with *-mekwisha-* treated as an MTA prefix is sometimes treated as a variant of the construction in (5.5) analysed as the auxiliary verb *kwisha* with the MTA prefix *-me-* but followed by a bare verb stem (i.e. minus the non-finite *ku-* prefix). The following transcription from Maw & Kelly (1975:110-1) illustrates this practice:

- (5.8) Je, ndio wakati huu u + me + kwisha timiza mihadi yenu,  
*Je really time that SP PERF kwisha fulfil promise your (pl)*  
 au u + me + kwisha vunja mihadi yenu?  
 or SP PERF *kwisha* break promise your (pl)  
*So, by that time, had you already fulfilled your promise,*  
*or had you already broken it?*

In other auxiliary plus main verb constructions, however, only the non-finite form of the main verb with the *ku-* prefix is possible, so *-mekwisha-* in (5.6) is best analysed as a MTA marker prefixed to a verb stem.

The fact that *-mekwisha-* has developed into *-mesha-* provides an insight into the typical prosody associated with utterances of constructions like (5.5) with a grammatical functional load, and lends support to Givón's iconicity principle, that intonation contours of serial verb constructions reflect cognitive packaging (such as whether or not the auxiliary verbal group is being used as a grammatical marker). Stress in Kiswahili, manifested by relatively high pitch, typically falls on the penultimate syllable of most (polysyllabic) words (stress is indicated by an accent over the stressed vowel):

- (5.9) Ni + me + kwísha     *I have finished*  
 SP PERF finish

The fact that this stressed syllable has been lost in the reduced form of the grammatical marker *-mesha-* suggests that, when used grammatically, serial verb constructions such as (5.5) function prosodically as a single verbal group with stress on the penultimate syllable of the main verb only:

- (5.5) A + me + kwisha     ku + sóma     *She has already read*  
 SP PERF *kwisha*     INF read

In their transcription of (5.8), Maw & Kelly (1975) observed that the syllable /sha/ functioned as what they term a 'salient syllable'. Maw & Kelly (1975) posit the 'tone-group' as the unit of the intonational system, identified by reference to a 'tonic' - a stressed syllable exhibiting one of a set of large-scale pitch movements. Salients initiate a pitch movement culminating in a tonic, and the beginning of a salient coincides with an accented syllable of the utterance (Maw & Kelly 1975:7). This entails that the syllable /sha/ was stressed in the utterance of (5.8) rather than /kwi/ since /sha/ marked the beginning of a salient. From a discourse perspective, stressing /sha/ emphasised what I have rendered as *already* in the translation of (5.8).

Once formal grammaticization has taken place and *-mekwisha-* is realised as a prefix on the verb, it behaves syntactically and morphologically as a typical MTA marker. Two co-occurrence phenomena illustrate this. First, all MTA markers, including *-mekwisha-*, can co-occur with the object prefix:

- (5.10) "U + mekwisha + mw + ona     boy     wa     somo     yako?" (RM p.34)  
 SP *mekwisha*     OP     see boy     from     class     your  
*Have you seen the boy from your class yet?*

Secondly, *-mekwisha-* has become one of a subset of Kiswahili MTA markers which can function as aspectual markers on the main verb of complex VPs. Complex VPs consist of a tensed form of the verb *kuwa* ('be') followed (not necessarily immediately) by the main verb with an MTA prefix marking aspect drawn from the set: *-na-* (5.11), *-ki-* (5.12), *-me-* (5.13), *-mekwisha-* (5.14) (the *-mekwisha-* prefix is more common in the main clause of complex VPs than *-me-*) and the negative forms *ha-ja-* (5.15) and *ha-i* (5.16), these latter two consisting of an initial prefix *ha-* plus a pre-stem prefix *-ja-* and a verbal suffix *-i* respectively.<sup>55</sup> The subject of the auxiliary

<sup>55</sup> Hauner (1984:112) also suggests that the prefix *-a-* is used as an aspect marker; however as all examples of this prefix occur in the first person singular (SP1, *ni*) e.g. *nilikuwa nataka*, the most plausible explanation is that the supposed *na-* (*ni+a-*) is in fact a contraction of *ni+na-*.

is the same as that of the main verb (in examples below *ha-* consists of *ha + a-*). Although any MTA prefix may in principle occur in the auxiliary, in practice *-li-* (past) and *-ta-* (future) predominate; in particular *-li-* is found in contemporary narrative, as in the following examples (all from RM p.6):<sup>56</sup>

- (5.11) Tangu Ijumaa mpaka Jumapili Zakaria a+li+kuwa  
 from Friday until Sunday Zakaria SP PAST be  
 a+na+kwenda kunywa...  
 SP PROG go drinking  
*From Friday to Sunday Zakaria went drinking...*
- (5.12) Wakati watoto wa+li+po+kuwa wa+ki+imba,  
 time children SP PAST REL be SP PROG sing  
 Zakaria a+li+kuwa a+ki+cheza - a+li+kuwa a+ki+rukaruka.  
 Zakaria SP PAST be SP PROG dance SP PAST be SP PRO jump-jump  
*Whilst the children were singing, Zakaria was dancing -  
 he was jumping up and down.*
- (5.13) Zakaria a+li+kuwa a+me+shindwa hata kujenga nyumba ya maana.  
 Zakaria SP PAST be SP PERF defeated even to build house of import  
*Zakaria was incapable even of building a decent house*
- (5.14) Kwa wakati huo wawili wa+li+kuwa wa+mekwisha+uzwa.  
 By time that two SP PAST be SP mekwisha be sold  
*By that time two (of them) had already been sold.*
- (5.15) ... a+li+kuwa ha+ja+toa hata chapa ku+wa+lipia  
 SP PAST be NEG-SP YET put out even stroke to OP pay for  
 watoto ada ya shule.  
 children fees of school  
*... he had not done even a stroke (of work) to pay for his children's  
 school fees.*
- (5.16) Stella a+li+kuwa ha+wez+i ku+ficha siri...  
 Stella SP PAST be NEG-SP be able NEG to keep secret  
*Stella was unable to keep a secret...*

---

<sup>56</sup> As examples (5.11) and (5.12) illustrate, *-na-* and *-ki-* convey similar information. According to Nurse & Hinnebusch (1993:381) *-na-* marks progressive aspect whilst *-ki-* conveys a more general imperfective meaning:

- (7) tu - li - kuwa tu - na - cheza 'we were playing (when you arrived)'  
 we PAST be we PROG play
- (8) tu - li - kuwa tu - ki - cheza 'we were playing (all day long, regularly)'  
 we PAST be we IPFV play or 'we used to play'

However, this analysis is clearly contradicted by our examples (5.11) and (5.12) in which the interpretations are reversed.





## 5.2 The causes of grammaticization

### 5.2.0 Introduction

From the existence of expressions which behave semantically like grams but formally as members of lexical categories, such as French *devoir*, and conversely from the non-existence of formally grammaticized expressions with full lexical semantics, combined with Givón's (1991) observations that functional and formal features of grammaticization do not always develop in parallel, it is clear that semantic change drives formal change, rather than being driven by it. This has generally been assumed throughout the literature on grammaticization (cf. Bybee *et al* 1994:281-301).

In terms of the conceptual/procedural distinction, the semantic change driving grammaticization (taken as a composite functional and formal development) is the addition of procedural information to the semantics of an expression, alongside the conceptual information already encoded. That is, a given expression not only encodes information giving access to its own logical selection frame but also comes to encode information giving rise to a logical selection frame other than that in which it appears. In other words it comes to constrain the interpretation of the associated utterance in some way. We saw in §2.1.1 that there is in principle nothing to stop a given expression encoding both conceptual and procedural information; the question to be answered here, however, concerns what causes lexical expressions encoding conceptual information to come to encode procedural information as well.

Bybee *et al* (1994:282) identify five mechanisms of semantic change involved in the development of grams marking tense, aspect and modality, some characteristic of early stages of the grammaticization process, others of later stages. Given that the shift from conceptual to procedural encoding is the driving force behind the formal changes involved in grammaticization, I shall concentrate on describing those mechanisms operative in the early stages of grammaticization with a view to establishing which can initiate the shift to procedural encoding. The five mechanisms of semantic change to be discussed are 'metaphorical extension', the conventionalisation of implicature,<sup>57</sup> which I shall generally refer to simply as 'inference' (following Bybee *et al* 1994), 'generalisation', 'harmony' and 'absorption' (of features of the linguistic contexts in which a gram prototypically occurs). I shall briefly describe each before looking in more detail at those involved in the early stages of grammaticization.

---

<sup>57</sup> The conventionalisation of implicature should not be confused with the Gricean notion of conventional implicatures, which I reject (see §1.3.3 and 2.1.1).

### 5.2.1 Metaphorical extension and inference

Metaphorical extension involves an abrupt shift of meaning from one, usually concrete, semantic category to another, usually more abstract one, for example, from OBJECT to SPACE to TIME to QUALITY. These can be viewed as primitive ontological categories similar to those proposed by Jackendoff (1983) (see §2.1.2 for discussion). Heine *et al* (1991:161) illustrate the effect of metaphorical extension on the use of the Ewe lexeme *megbé* ('back'); in (5.19) *megbé* denotes a body part (OBJECT), in (5.20) a location (SPACE) either as an adverb (5.20a) or as a postposition (5.20b), in (5.21) a temporal relation (TIME) and finally in (5.22) it means 'mentally retarded' (QUALITY):

					<i>semantic category</i>	
(5.19)	é-pé		megbé	fá	OBJECT	
	3SG-POSS		back	be cold		
	<i>His back is cold</i>					
(5.20) a.	é	le	xɔ	á	megbé	SPACE
	3SG	is	house	DEF	behind	
	<i>He is at the back of the house</i>					
	b.	é	nɔ		megbé	SPACE
		3SG	stay		behind	
	<i>He stays back</i>					
(5.21)	é	kú	le	é-megbé	TIME	
	3SG	die	be	3SG-behind		
	<i>He died after him</i>					
(5.22)	é	tsi	megbé		QUALITY	
	3SG	remain	behind			
	<i>He is backward/mentally retarded</i>					

(5.22) is ambiguous between 'He is mentally retarded' and 'He remained behind' or 'He is late'; this, and similar ambiguity between SPACE and TIME uses and OBJECT and SPACE uses, reflect some overlap between consecutive semantic categories.

Despite the attention paid to metaphorical extension as a potential cause of grammaticization, Bybee *et al* (1994:283) observed that very few examples of grammaticization of tense, aspect or modality in their study could be accommodated in the metaphorical model, and of those that could, the change occurred very early in the grammaticization process, exhibiting features of lexical rather than of grammatical change. A similar observation was made by Traugott (1988) who argued that although ordinary semantic change was metaphoric, grammaticization involves a special kind of metaphorical extension: metonymy, or the conventionalisation of implicature (see footnote 57). Carey (1990) who describes the role of the conventionalisation of implicature (henceforth 'inference') in the early development of the English perfect, characterises this mechanism as follows:

"When semantic change arises from the conventionalisation of invited inferences, some aspect of the context in which the expression was used with its old meaning becomes indexed and over time becomes part of the new meaning of the expression itself."

(Carey 1990:373)

### 5.2.2 Generalisation, harmony and absorption

Three further mechanisms of semantic change are discussed in Bybee *et al* (1994), the first of which is generalisation, "the loss of specific features of meaning with the consequent expansion of appropriate contexts of use for a gram" (Bybee *et al* 1994:289). Generalisation can be described in terms of the loss of one feature of meaning; this is neatly illustrated in the development of *can* from being a marker of purely mental ability to general ability and finally to representing root possibility:

*can*

- |       |                  |  |
|-------|------------------|--|
| (i)   | mental ability   | mental enabling conditions exist in an agent for the completion of the predicate situation |
| (ii)  | general ability  | enabling conditions exist in an agent for the completion of the predicate situation        |
| (iii) | root possibility | enabling conditions exist for the completion of the predicate situation                    |

(Bybee *et al* 1994:290)

Generalisation may also be the mechanism behind the development of negation markers such as English *nought* > *not* and French *ne...pas*, both of which were formerly emphatic negation markers but which have now been generalised to non-emphatic contexts (Bybee *et al* 1994:293).

The final two mechanisms of semantic change to be discussed are harmony and the absorption of contextual meaning. Where harmony occurs, the meaning of a gram does not contribute any new information to a sentence since it merely echoes the temporal reference, modality, etc. expressed in the main clause (for example through temporal or modal adverbials). Harmony applies only to the later stages of grammaticization (Bybee *et al* 1994:294) and results in semantic redundancy. A gram can come to be used so frequently, including when it harmonises with the context and is semantically redundant, that its absence becomes significant. When this occurs, the absence of such a gram may result in the formation of what Bybee *et al* (1994:294) call a 'zero gram', which absorbs contextual meaning pertaining to the domain constrained by the overt gram. For example, the absence of the past tense *-ed* marker signals unmarked tense (not present tense as Bybee *et al* (1994:294) claim) whereby

the addressee recovers the intended temporal interpretation from the context and the absence of an overt *-ed* marker (see §3.3.2 for discussion).

These latter two mechanisms - harmony and the absorption of contextual meaning - are only involved in the later stages of grammaticization and so will not be discussed further. Instead, I shall focus on metaphorical extension, inference, and generalisation, and attempt to determine which of these is/are involved in the shift from purely conceptual encoding to procedural encoding and grammaticization.

### 5.2.3 The mechanism of semantic change in grammaticization

Heine *et al* (1991) claim that metaphorical extension is one of the main, if not the main, mechanisms of semantic change involved in grammaticization. However there are two strong arguments against this view. The first is that the clearest cases of metaphorical extension, such as the extension of body-part terms to express spatial concepts (for example the extension of the semantic field of the Ewe lexeme *meḡbé* discussed above), are arguably not instances of grammaticization (cf. Bybee *et al* 1994:284). At best, metaphorical extension often precedes subsequent grammaticization and may be a prerequisite for it. Secondly, the clearest cases of grammaticization are arguably due to inference rather than to metaphorical extension. These two mechanisms can be difficult to distinguish, given that both can be viewed as ways of solving the problem of how to represent members of one semantic category in terms of another (Traugott 1988:413, Heine *et al* 1991:150). Traugott (1988) defines metaphor as the representation of information pertaining to one semantic domain in terms of another semantic domain not present in the context, and metonymy (inference) as the representation of information pertaining to one semantic domain in terms of another semantic domain which is present in the context, albeit non-overtly.

The difficulty in distinguishing between metaphorical extension and inference can be illustrated through the development of an epistemic interpretation of *must* from the historically prior obligation interpretation. Bybee *et al* (1994:284) argue that this change is due to metaphorical extension since "the obligation and probability senses of *must* occur in mutually exclusive environments." The obligation reading occurs in future contexts such as (5.23) and the probability reading in past or present contexts such as (5.24):

(5.23)           The letter must arrive sometime next week.

(5.24)           The letter must have been in the mail.

Since, Bybee *et al* reason, an epistemic interpretation is not available in (5.23) and a deontic interpretation is not possible in (5.24), the extension of meaning from the deontic to the epistemic semantic field is a shift from one category to another not present in the context, and must therefore be a case of metaphorical extension. However, it is possible to find contexts in which examples such as (5.23) can receive an epistemic interpretation:

(5.23')            With a first class stamp it shouldn't take more than three days to arrive.  
                      The letter **must** arrive sometime next week.

The epistemic interpretation of *must* in (5.23') derives ultimately from the fact that the postal services are obliged to convey a letter with a first class stamp to its destination within a certain time limit. The epistemic interpretation is therefore based on a semantic category (obligation) which is present in the context, hence the mechanism at work is inference rather than metaphor. Similarly, the epistemic interpretation of (5.24) can be viewed as deriving from obligation, given suitable contextual support (e.g. the secretary was obliged to follow the instruction to mail the letter):

(5.24')            I told the secretary to mail the details last week,  
                      so the letter **must** have been in the mail.

These examples are not intended as knock down arguments against Bybee *et al* (1994:284), but simply aim to demonstrate the difficulty of distinguishing metaphor and inference.

Turning now to inference as a mechanism of semantic change, Traugott (1988:413) characterises it as crucially involving "explicit coding of relevance and informativeness that earlier was only covertly implied." In other words, grammaticization involves the conventionalisation of implicatures, which can be viewed as the licensing of inferences involved in uncovering 'covert' (i.e. implied) meaning, even in the absence of direct contextual support. Thus, for example, from an utterance of (5.23') an addressee could infer - from the fact that the postal services have certain obligations - that the letter in question probably will arrive the following week. With increased grammaticization, the inference to an epistemic interpretation can be made even in the absence of a deontic statement from which epistemic probability could be implied.

Procedural encoding constrains the inferential processes involved in utterance interpretation in just this way, suggesting that inference is at least one mechanism of

semantic change which results in a shift from purely conceptual encoding to procedural encoding also. Inference, or the conventionalisation of implicature, can be illustrated in greater detail with respect to the development of *will* and *shall* (§6.2 discusses *will* and *shall* synchronically in greater detail).

Bybee *et al* (1994:287-8) describe the development of *will* and *shall* as grams in terms of the following three stages:

Use with first person subjects (obligation with *shall*, desire with *will*)

(INFERENCE) ---> intention with first person subjects

(INFERENCE) ---> prediction

This development can be described in more detail as follows, taking into account the effect of generalisation and a number of issues to be discussed in chapter 6:

Use with first person subjects (obligation with *shall*, desire with *will*)

(INFERENCE) ---> intention with first person subjects

(INFERENCE) ---> prediction about future events with first person subjects

(GENERALISATION) ---> prediction about future events (any subject)

(GENERALISATION) ---> prediction in general (*will* only)

Here generalisation is involved in the later stages of the grammaticization process, after the shift to procedural encoding has taken place. Generalisation also seems to be involved in the early semantic development of the *be going to* construction. The change from progressing physically towards a goal ('I am going to the shops') to progressing in a non-physical sense towards a goal, giving rise to the overtones of intention ('I am going to go shopping') and inevitability ('I am going to be sick') can be viewed as due to generalisation (the loss of the physical component of meaning). In §6.4 I shall demonstrate that the overtones of intention and inevitability often associated with the use of *be going to* are conceptual rather than procedural. I have found no case of generalisation changing a lexical to a grammatical expression (i.e. initiating procedural encoding), suggesting that generalisation as a mechanism of semantic change is not itself the mechanism by which grammaticization is activated through the shift to procedural encoding, although it can derive a new, more general lexical meaning from a lexical expression, and a more general grammatical meaning from a gram. That leaves just inference, or the conventionalisation of implicature, as the one sure mechanism of semantic change driving the shift to procedural encoding which initiates grammaticization.

## 5.3 Semantic retention

### 5.3.0 Introduction

I have so far argued that grammaticization is semantically instantaneous but formally gradual, and that the mechanism of semantic change which triggers grammaticization is the conventionalisation of implicature, giving rise to procedural encoding at phase two of utterance interpretation. That grammaticization is semantically instantaneous is possible because expressions may encode both conceptual and procedural information. Any procedural information encoded by an expression is automatically recovered (in addition to being merely accessed/activated) since it provides a necessary processing constraint on the interpretation of an associated conceptual representation. If the resulting interpretation is relevant on its own, the interpretation process will cease so as to minimise processing effort. However, when a formerly lexical expression is used as a grammatical marker, it does not suddenly cease to encode conceptual information; this conceptual information may no longer be of prime importance to the interpretation of an utterance containing such an expression, but it is nonetheless still accessible (that is, activated during modular decoding).

As a gram develops over time, the conceptual information it encodes may become increasingly inaccessible, until such time as it disappears altogether and the gram encodes only procedural information, as does the English past tense suffix, *-ed*. However, from the moment an expression first encodes procedural information until the point where, as a gram, it no longer encodes any conceptual content, both conceptual and procedural information are encoded and may be recovered. Throughout the period of its development in which a gram encodes both conceptual and procedural information, it is said to exhibit semantic retention, whereby "certain more specific semantic nuances of the source construction can be retained in certain contexts long after grammaticization has begun." (Bybee *et al* 1994:16)

Synchronic semantics since Saussure has avoided recourse to etymology in describing the semantic content of any expression, and the semantics of grammatical markers is no exception. As Culler (1976:38) notes, knowing that the French negative marker *pas* and the noun *pas* ('step', 'pace') derive from the same source is irrelevant as far as a synchronic account of the French negative marker is concerned. Nonetheless, certain semantic features of grams, such as the volitional interpretation of *will* which may be recovered in interrogative clauses, and also in certain *if*-clauses containing a variant of *will*, can only be accounted for in the light of the historical development of the gram in question; that is, as examples of semantic retention. In equating semantic retention with residual conceptual encoding we can maintain a monosemous



synchronic characterisation of grams such as *will*, but at two distinct levels of information structure: conceptual and procedural. After briefly discussing the problems caused by semantic retention for traditional polysemous and monosemous semantics, I shall further discuss the case of semantic retention in *will*.

### 5.3.1 Problems for traditional analyses

The term 'semantic retention' is something of a misnomer - grams which do not exhibit 'semantic retention' nonetheless have semantic content, albeit purely procedural; others, however, retain a certain amount of conceptual semantic content which is accessible in certain contexts. Hence 'conceptual retention' might be a more accurate term. Nonetheless, in the spirit of Occam's Razor (terminology is not to be increased beyond necessity) I shall adopt the term 'semantic retention' as used in Bybee *et al* (1994:15-18). The term 'persistence' has also been used (Hopper 1991:28-30) but this reflects a polysemous view in which grammatical markers exhibiting 'persistence' have two meanings: one lexical and one relational. The term 'semantic retention', however, carries no such connotations. Bybee *et al* (1994:281) profess agnosticism over whether grams such as *will* are synchronically monosemous or polysemous, and Bybee (1988:255) explicitly rejects Coates' (1983) analysis of *be going to* as polysemous. This is compatible with the monosemous, unitary semantic approach adopted here. Semantic retention is problematic for polysemous accounts of grams, in that the specific semantic nuances of the source construction which distinguish, say, a supposed volitional *will* from a purely grammaticized *will* are often difficult to isolate.

(5.25) I'll do the shopping.

(5.25) could be loosely paraphrased as "I am willing to do the shopping" or could be interpreted as a simple prediction; often, however, (5.25) will be interpreted as a prediction with overtones of volition. On a polysemous account, we would have to say that *will* here is ambiguous between two meanings, or means two things at once.

If *will* is characterised polysemously, any utterance containing it will need to be disambiguated; if *will* is characterised (as in §6.1 and §6.2) as monosemous and semantically underdetermined, it will need to be inferentially enriched. Either way, inferential processes must be brought to bear on the interpretation process. On the basis of Modified Occam's Razor, however, monosemous accounts are to be preferred wherever possible. However, semantic retention is also problematic for traditional monosemous accounts, which characterise grams as pure relational elements (Bybee

*et al* 1994:17). Restrictions on the distribution of grams and differences in their interpretation can often be accommodated within a monosemous account by reference to an underlying unitary semantics for each gram, combined with an account of the systematic relations existing between grams sharing a common semantic field (such as was proposed for *may*, *can*, *must* and *should* in §4.2). However, semantic retention, such as the volitional overtone of *will*, is of a different order from the epistemic, deontic and dynamic interpretations of other modal auxiliaries. The latter can be derived from a combination of some single underdetermined procedural semantics for a given modal, plus an account of the systematic interaction of the modal auxiliaries with each other and with the principle of relevance. The volitional overtones of *will*, on the other hand, can not be derived in this way; chapter 6 demonstrates that the procedural semantic ontology of *will* alone is too indeterminate to give rise a specifically volitional interpretation, and that the systematic variation of *will* with other grams (notably *shall* and *be going to*), while it gives rise to certain implicatures, does not give rise to overtones of volition.

### 5.3.2 Semantic retention: the case of *will*

This chapter concludes with an explanation of semantic retention within the model of grammaticization proposed here. I shall continue to use *will* for exemplification, for the following reasons. First, because *will* provides a clear illustration of semantic retention, and secondly, to anticipate the more detailed discussion of *will*, *shall* and *be going to* in chapter 6. There are, however, two non-grammaticized forms of *will* which can not be accommodated within a monosemous account of grammaticized *will*.<sup>58</sup> I shall first distinguish these non-grammaticized expressions from the homophonous grammatical marker, so as to exclude them from my discussion of semantic retention with grammaticized *will*.

The first non-grammaticized expression is *will* with habitual dynamic modal interpretation. There are three main reasons for excluding this 'habitual *will*' from our discussion of the grammaticized form: First, the interpretation of habitual *will*, that the subject of the clause is in the habit of doing something (for example in the collocations "if you will keep on/insist on doing x..."), is not recoverable as an overtone of grammaticized *will*. Secondly, habitual *will* must be emphasised with the result that, unlike grammaticized *will*, it does not have a reduced allomorph [ɪ] affixed to the subject of the clause in which it occurs (compare (5.26b) and (5.26c) below). The negated form *won't* is possible as this can be independently emphasised (see

---

<sup>58</sup> Inflected, and often somewhat archaic, variants of *will* (e.g. "I must do as she wills", "He willed the horse to win") can be ignored.

(5.27) below). Third, habitual *will* is always possible in the protasis of conditional clauses<sup>59</sup> when the grammaticized form is generally not (compare (5.26b) and (5.26c) below):

- (5.26) a. If you go out in the rain, you'll get wet.  
b. \* If you'll go out in the rain, you'll get wet.  
c. If you will go out in the rain, you'll get wet.

(5.27) If you won't wrap up warm, of course you'll catch a chill.

The other non-grammaticized expression arises from a fossilised usage of "the earlier 'desire' meaning of the now-grammaticized *will*" Bybee *et al* (1994:16):

- (5.28) a. Do what you will, I won't betray my comrades. (Bybee *et al* 1994:17)

The volitional interpretation of *will* in (5.28a) is similar to that conveyed by cases of semantic retention with grammaticized *will*, but formally the two are in complementary distribution. In the purely volitional usage, *will* does not take a complement, either overt or inherent, and the reduced allomorph is ungrammatical, as evidenced by (5.28b):

- (5.28) b. \* Do what you'll, I won't betray my comrades.

The fact that, in the habitual and purely volitional examples, *will* has no phonetically reduced allomorph suggests a lexical rather than a grammaticized characterisation and so these forms of *will* may be excluded from any synchronic characterisation of the homophonous grammatical marker, which must take account of semantic retention. Semantic retention with *will* is most evident in two environments: in interrogative clauses with second person subjects, and in the protasis of certain conditional clauses. The reasons for the prevalence of semantic retention in these two environments differ; in interrogative clauses semantic retention results from the nature of questions in the second person, the contrast with other future expression, and possibly also the fact that the phonetically reduced allomorph 'll is not possible. In the protasis of conditional clauses, semantic retention arises when *will* is in harmony with the context set up by the conditional marker *if*. I shall begin by discussing *will* in interrogative clauses.

---

<sup>59</sup> See below for further discussion.

### Will in interrogative clauses

Bybee *et al* (1994:16) suggest that some of the differences in interpretation between *will*, *shall* and *be going to* are the result of semantic retention. Comparing *will* and *shall*, they claim that *shall* is appropriate in first person questions because its “obligation sense” suggests a request for confirmation from the addressee of the speaker’s responsibility,<sup>60</sup> whereas *will* is not appropriate because “the older ‘desire’ sense of *will* seems to come through, suggesting that the speaker is quite inappropriately asking whether s/he wants to call a cab”:

- (5.29) a. Shall I call you a cab?  
b. Will I call you a cab?

The volitional overtones of *will* are also highly accessible in interrogative clauses in the second person, which are frequently interpreted as requests, i.e. questions about the addressee's willingness to perform a particular function, from which a request to perform that function can be inferred. There are a number of factors which may contribute to this.

First, requests in English are often expressed through the use of indirect illocution, frequently employing interrogatives. In a context where interpreting a second person interrogative with *will* simply as a request for information about a future situation or event fails to yield adequate contextual effects, an indirect illocution may be recovered, based on an interpretation of the interrogative as a request for information about the addressee’s willingness to perform a particular act. The use of *will* rather than other expressions to make requests reflects the fact that volitional overtones are more accessible from the use of *will*, which has developed from a volitional source expression, than from the ‘present progressive’ or *be going to*, for example, which have not. Whilst interrogatives with *will*, the ‘present progressive’ and *be going to* can all be used to request information, only *will* can be used to make a request:

- (5.30) a. Will you give tomorrow’s lecture, or don’t you know yet?  
b. Are you giving tomorrow’s lecture, or don’t you know yet?  
c. Are you going to give tomorrow’s lecture, or don’t you know yet?
- (5.31) a. Will you give tomorrow’s lecture please?  
b. ? Are you giving tomorrow’s lecture please?  
c. ? Are you going to give tomorrow’s lecture please?

---

<sup>60</sup> In §6.2 I shall argue that *shall* does not in fact exhibit semantic retention, and that some of the differences in acceptability and interpretation between *will* and *shall* are due to differences in the procedural information these expressions encode.

Secondly, in interrogative clauses, only the phonetically full form of *will* is possible, although this need not be stressed at all:

- (5.32) a. Will you do the shopping?  
b. \* 'll you do the shopping?

Recall that neither of the non-grammaticalized forms of *will* encoded procedural information and neither had a reduced allomorph 'll. According to the model of conceptual and procedural encoding developed here, these two facts - the semantic and the formal - are connected. Conceptual information is information encoded by an expression which gives access to its own logical selection frame; that is, conceptual encoding gives rise to a concept capable of combining with other concepts to form a logical form. One concept within a logical form can be highlighted by emphasising the expression encoding it; conversely, an expression encoding a less important concept may receive less emphasis. Formal grammaticization often leads to a loss of conceptual semantic content since, as information encoded by an expression giving access to its own logical selection frame becomes less important the expression becomes formally less conspicuous, which in turn reduces the salience of the conceptual information and so on until all conceptual information has been 'bleached' from the expression. However, when a gram is highly salient, for example when it occurs in its full phonetic form and at a prominent point in a clause (for example at the beginning), and in a context which invites an appropriate interpretation, the conceptual information it encodes also becomes more highly salient; this, then, is one factor contributing to semantic retention.

In comparison to *will*, take the Kiswahili future marker *-ta-*. All MTA markers in Kiswahili occur as prefixes to the verb stem. As we saw in §5.1.3, stress in Kiswahili typically falls on the penultimate syllable of a word or phrase, and most MTA markers (including *-ta-*) never fall on the penultimate syllable of a verb phrase; if the verb stem is monosyllabic the infinitival prefix *ku-* is inserted between the MTA prefix *-ta-* and the verb stem and takes stress (indicated by an accent):

(5.33) ni + ta + kwénda  
1S FUT go  
*I will go*

(5.34) ni + ta + kú + ja  
1S FUT INF come  
*I will come*

Questions in Kiswahili do not involve an interrogative word order, instead declarative word order is retained and the interrogative function is indicated through *je* and/or intonation (see §2.2.3), so again, the MTA prefix cannot take stress:

- (5.35) "Je, Najum, utakwénda sokóni?"  
*Je, Najum, you will go to the market?*

Correlating with this, *-ta-* does not exhibit semantic retention, even though it shares a similar lexical source to *will* in the verb *taka* ('want', or less frequently 'be about to') which is a productive, regular verb in contemporary Kiswahili. Also contributing to the interpretation of example (5.35) as a question rather than as a request is the fact that indirect illocution is far less common in Kiswahili than in English; if the speaker wanted Najum to go to the market, he would generally tell him so directly.

#### *Will in conditional clauses*

Semantic retention also occurs with *will* in the protasis of certain conditional clauses. In a sentence of the form 'if *p* then *q*', *will* is generally not possible in *p*; instead, an unmarked tense form of the verb is generally used:

- (5.36) a. \* If it'll rain, you'll get wet.  
b. If it rains, you'll get wet.

This is an example of temporal subordination. Wekker (1980) and Declerck (1984) both propose that temporal subordination of subordinate to main clauses results in the deletion of an underlying future tense (*will*) in the protasis. Although they are incorrect to view *will* as a future tense (see §6.1) this argument can be applied as it stands to languages such as French which do use a future tense in the main clause of a conditional sentence but not in the subordinate clause. If futurity in English is inferred rather than encoded (as argued in §6.1) then the argument for temporal subordination still holds on the assumption that future time reference is inferred with respect to the whole conjunct, rather than with respect to each clause separately. Declerck (1984:283-4) argues that temporal subordination only occurs in sentences of the form 'if *p* then *q*' where *p* is a sufficient condition of *q*, *p* and *q* are temporally related with respect to each other, and both *p* and *q* express the same kind of condition ('open', hypothetical or counterfactual).<sup>61</sup>

---

<sup>61</sup> cf. Depraetere (1994) and Declerck & Depraetere (1995) for discussion. (In addition, Haegeman & Wekker (1984) demonstrate that conditional clauses with temporal subordination differ syntactically from those without temporal subordination.)

When these conditions hold and the apodosis has past time reference, temporal subordination does not result in the loss of the past tense *-ed* suffix in the protasis in favour of the unmarked form:

- (5.37) a. If you answer/\*will answer that question, you will win a prize.  
b. If you answered that question, you would win a prize.  
c. If you had answered that question, you would have won a prize.

[based on Declerck 1984 (8a) to (8c)]

Unlike *will*, *-ed* can exhibit harmony with the context (see §5.2.2). When the procedural information encoded by *will* is in harmony with the context, the occurrence of *will* is still marked in comparison to the unmarked tense form and addressees will therefore look for additional contextual effects to justify the additional processing effort necessitated by the use of the relatively more marked *will*. Unlike the *-ed* suffix which has been bleached of all conceptual content, *will* still encodes conceptual information relating to volition; this conceptual information becomes accessible through the principle of relevance (the additional processing effort incurred by the use of *will* + lexical verb over the unmarked form of the lexical verb initiates in a search for additional contextual effects, which results in the conceptual information encoded by *will* being recovered).

If the conceptual information encoded by *will* is semantically incompatible with the subject of the protasis, the resulting sentence is deemed ungrammatical, as in (5.36) repeated below:

- (5.36) a. \* If it'll rain, you'll get wet.  
b. If it rains, you'll get wet.

Where the subject of the protasis is semantically compatible, the conceptual information encoded by *will* becomes accessible. This is illustrated in the following examples, from Comrie (1985:120):

- (5.38) a. If you do the shopping for me, I'll give you some money.  
b. If you'll do the shopping for me, I'll give you some money.

(5.38b), in which the shopping occurs after money has been exchanged, could be loosely paraphrased by (5.39) with the same relative temporal ordering:

- (5.39) If you are willing to do the shopping for me, I'll give you some money.

In (5.38b) *p* ('you'll do the shopping for me') and *q* ('I'll give you some money') fulfil all the criteria for temporal subordination, as in (5.38a): *p* is a sufficient condition of *q*, *p* and *q* are temporally related with respect to each other, and both *p* and *q* express the same kind of condition ('open'). In other words, the procedural information encoded by the *will* in the protasis of (5.38b) is in harmony with the context, but *will* occurs nonetheless. As (5.38a) demonstrates, *will* could be replaced here with the unmarked tense form of *do*, but (5.38a) and (5.38b) are not synonymous; only (5.38b) can be paraphrased by (5.39). This is because of semantic retention, in which the conceptual information encoded by *will* is recovered by the addressee on the basis of the principle of relevance.

Because of the possibility of paraphrasing *will* in sentences such as (5.38b) with *be willing to* as in (5.39), Coates (1983:172) characterised *will* as primarily expressing volition.<sup>62</sup> Against this view, Klinge (1993:344) claims that, when *will* seems to assign volition to the referent of its subject NP, as in (5.40), this is in fact a property of the utterance rather than a case of semantic synonymy between *will* and genuinely volitional forms such as *be willing to* (5.41) or *want to* (5.42), as the unacceptability of (5.43) demonstrates:

- (5.40) I will help you.
- (5.41) I am willing to help you, but unfortunately I cannot.
- (5.42) I want to help you, but unfortunately I cannot.
- (5.43) ! I will help you, but unfortunately I cannot.

[Klinge 1993 (69) to (72)]

Now, Klinge (1993) claims that any volitional interpretation of an utterance of a sentence containing *will* is inferentially inferred from the utterance interpretation context rather than semantically encoded by *will*. This does not explain why volition in particular is so frequently inferred from the use of *will*, if nothing in the semantics of *will* refers to volition. However, if we claim, as I have done, that *will* encodes conceptual information relating to volition, and that this is only recovered when licensed by the principle of relevance, the fact that volition is often, but not always, recovered can be explained. Whilst the conceptual information encoded by *will* is not always recovered (although it is 'accessed' in the psycholinguistic sense), the procedural information it encodes (informally defined as 'prediction', but see §6.1) is. The unacceptability of (5.43) arises because the procedural information encoded by

---

<sup>62</sup> Barese (1980:51) also treats *will* as either a future tense, or as "short for some form of 'to be willing'," which is not future but rather signals present intention (the 'intentional' reading).



*will* (that the SR [I HELP YOU] is predicted) contradicts the second clause ('I cannot help you' in which [I HELP YOU] is not predicted).

#### **5.4 Summary**

In chapters 3 and 4 I demonstrated that grammatical markers of tense, aspect and modality are exponents of procedural encoding at phase two of utterance interpretation. They constrain the inferential interpretation of sub-propositional conceptual representations of situations and events (SRs) with respect to features of the utterance interpretation context, and were termed 'operators'. Given that operators are grammatical markers ('grams'), a comprehensive account of their functions and distribution must take into account the fact that grams develop from lexical sources and may still be developing. That is, the procedural account of operators developed in chapter 3 must be compatible with what is known about grammaticization.

In this section I have demonstrated that a procedural account of grammatical markers is compatible with research into grammaticization only if we allow linguistic expressions the capacity to encode both conceptual and procedural information. This modification, the possibility of which was demonstrated theoretically in §2.1.2 has a number of consequences for the relation between grammaticization and the conceptual/procedural distinction.

Because it is a formally gradual process, grammaticization had been viewed as semantically gradual also. On this view the onset of grammaticization and the subsequent development of an expression from lexical to grammatical are all gradual, with semantic changes occurring parallel to formal changes such as phonetic reduction and affixation. This view is at odds with the clear-cut binary division of linguistically encoded information types into conceptual and procedural. Given the model developed in this thesis, we should expect a grammaticizing expression to encode both conceptual and procedural information, and that the onset of procedural encoding would be instantaneous. This is what Givón's (1991) study of phonetic iconicity (through pause length) with respect to serial-verbs in fact found. The characterisation of grammaticization as semantically instantaneous but formally gradual also goes some way to explaining the existence of formally lexical verbs with clearly grammatical semantic content, such as modal verbs in French, German and Spanish.

In §5.2 I demonstrated that the conventionalisation of implicature appears to be the only sure mechanism of semantic change underlying the onset of procedural encoding

(i.e. initiating grammaticization). Given that the purpose of procedural encoding is to constrain the inferences drawn from utterances, it is unsurprising that the conventionalisation of inference should give rise to procedural encoding.

Finally, the claim that a single expression may be an exponent of both conceptual and procedural information predicts the phenomenon of semantic retention. Semantic retention is at odds both with polysemous analyses of grams which posit ambiguity where there is none, and with monosemous accounts, based on a single underlying meaning for each gram, which fail to generate interpretations derived from semantic retention. Having disposed of variants of *will* in complementary syntactic and phonological variation with the homophonous grammatical marker, I demonstrated that semantic retention results from conceptual information encoded by *will* being made accessible to the addressee. This only occurs when *will* itself is sufficiently salient and co-occurs with appropriate agentive subjects. This may result when *will* is emphasised, as in interrogative clauses, or if the use of *will* in the protasis of a conditional clause is in harmony with the context, in which case the principle of relevance predicts that additional contextual effects will be sought.

In chapter 6 I shall analyse *will*, *shall* and *be going to* as exponents of procedural encoding at phase two of utterance interpretation (i.e. as operators), but taking into account the effects of semantic retention on the interpretations and distribution of these expressions. These expressions have been chosen because they represent different stages of grammaticization, exhibit semantic retention, and can be readily compared since they are all used in the expression of future time reference in English.

## 6 FUTURE EXPRESSIONS IN ENGLISH: *will*, *shall* and *be going to*

### 6.0 Introduction

Future time reference in English involves (not exclusively) the modal auxiliaries *will* and, to a lesser extent, *shall*, and the *be going to* construction. All three expressions are to be characterised as operators: exponents of procedural encoding at phase two of utterance interpretation with scope over SRs (although the scope of *be going to* is more restricted than that of the modal auxiliaries). Although procedural encoding is central to the semantics of *will*, *shall* and *be going to*, conceptual encoding must also be taken into account. Just as the ability of expressions to encode both conceptual and procedural information was shown to be central to the analysis of *will* in relation to grammaticization, so a synchronic characterisation of *be going to* (at least in contemporary British English) must make reference to both types of linguistically encoded information.

The conceptual information encoded by *will* was discussed in §5.3.2, so in this chapter I shall focus on the procedural information it encodes. *Will* can be viewed as the pivot of the current chapter; *shall* and the *be going to* construction will be analysed in comparison with it. *Shall* is an older gram than *will* (Bybee 1988:253), which appears to be being displaced by *will* and which, unlike *will*, does not exhibit semantic retention; *be going to*, in contrast, is a more recent grammatical construction than *will*, which appears to be increasing in frequency in comparison with *will* (at least in some dialects of English), and exhibits a higher degree of semantic retention than *will*.

In §6.1 I discuss the procedural semantics of *will* in both its future time and non-future time uses. In §6.2 *shall* is characterised in relation to *will*, and I shall conclude by discussing *will* and the *be going to* construction.

### 6.1 Will

#### 6.1.0 Introduction

In the previous chapter I used *will* as an example of an operator exhibiting semantic retention, in the form of overtones associated with its volitional source construction. Little was said concerning the procedural information encoded by *will*, so in this section I shall provide an account of the procedural semantics of *will*. In §6.1.1 I shall demonstrate that, unlike *can*, *may*, *must* and *should*, *will* does not relate SRs to assumptions in the utterance interpretation context, and can therefore be viewed as a

semantically atypical modal auxiliary. In §6.1.2 I shall discuss the relation between *will* and future time reference, and demonstrate that *will* does not relate SRs in its scope to the temporal cognitive domain, although this relation can be inferred from the use of *will* and the utterance interpretation context.

### 6.1.1 *Will* and modality

In line with my previous arguments (§4.1.2 and §5.3) in favour of monosemous characterisations of the modals, I shall maintain a monosemous perspective in describing the procedural information encoded by *will*. Although *will* behaves syntactically as a modal auxiliary, semantically it differs significantly from the members of this syntactic class discussed in chapter 4. The idea that *will* is an atypical modal auxiliary is not new:

"it seems that *will* makes a somewhat different contribution to the proposition expressed than the four modals discussed above. This intuition is shared by Sweetser (1989) who calls *will* and *shall* BORDERLINE modals, and is reflected in Ehrman's (1966) basic meaning for *will*, which she gives as: 'the occurrence of the predication is guaranteed', which can be interpreted as 'don't worry about the evidence, concentrate on the proposition that *will* modifies'."

(Groefsema 1995a:63)

This can be illustrated by comparing utterances with *will* and *must*. Traditionally, two types of 'modal certainty' have been recognised, one based on repeated experience or common sense and realised by *will*, which contrasts with 'logical necessity', based on deduction and realised by *must* (Downing & Locke 1992:385-6). In fact, *will* need not be restricted to modal certainty based on experience or common sense; compare (6.1) and (6.2):

(6.1) It will be very hot in Djibouti at the moment.

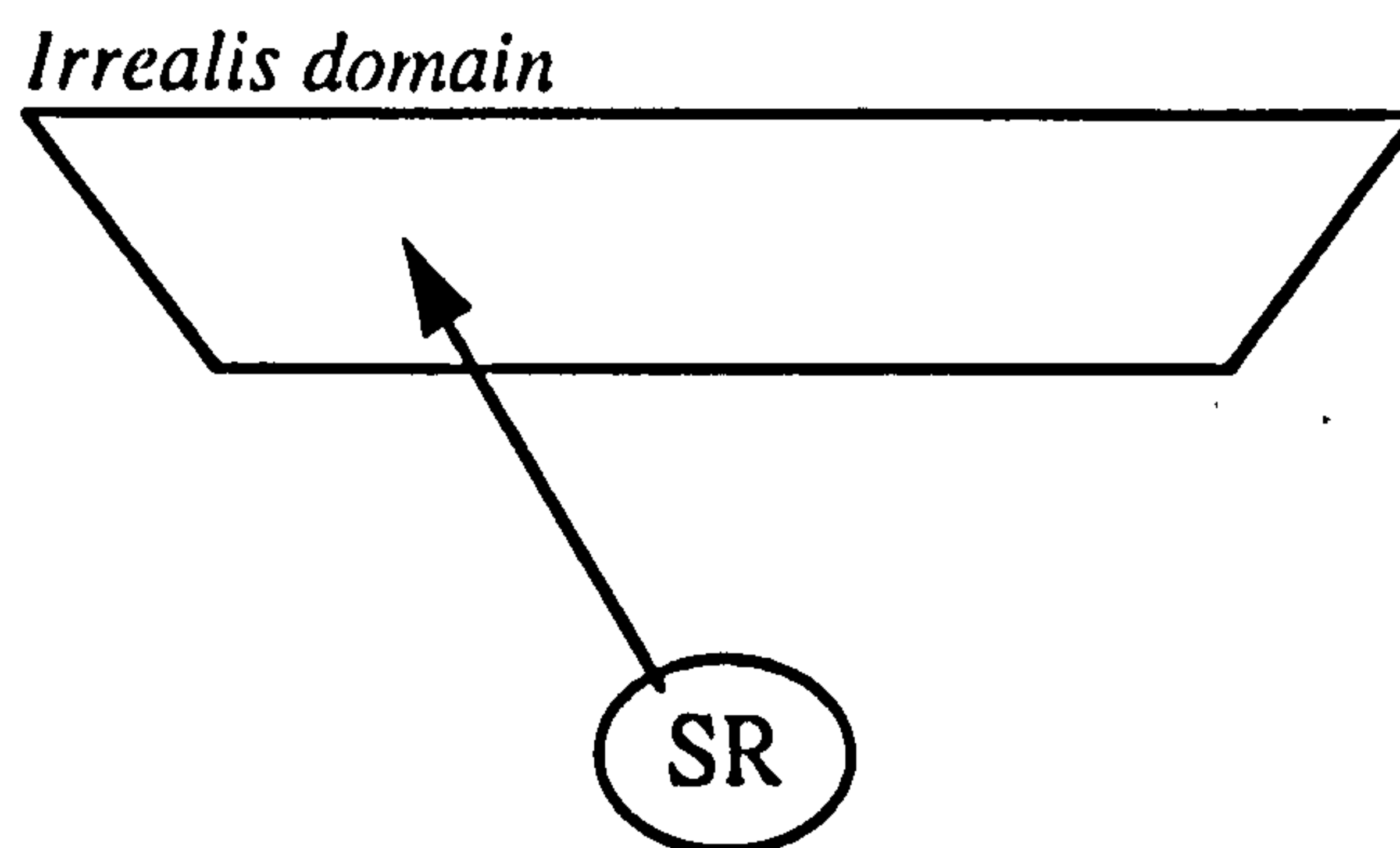
(6.2) It must be very hot in Djibouti at the moment.

Spoken in August by a person who has never been to Djibouti but who has access to some relevant assumptions (for example that Djibouti is the world's hottest capital city and that it is in the northern hemisphere) both (6.1) and (6.2) could be felicitous, although (6.2) with *must* can draw attention to the assumptions or the deductive process in a way which (6.1) with *will* does not. This is expected given that *must* encodes an instruction to treat any SR within its scope as entailed by all the assumptions that have a bearing on that SR. However, if I were to talk about the temperature in Djibouti whilst reminiscing about the year I spent working there, only (6.1) with *will* would be felicitous, since the proposition expressed would be based not on a deduction from relevant assumptions but on induction from prior experience.

This suggests that *will* does not constrain the interpretation of a SR within its scope with respect to those assumptions in the utterance interpretation context which constitute evidence for the occurrence of the SR. In this respect *will* differs from *can*, *may*, *must* and *should*. However, what both (6.1) and (6.2) have in common is that the SR [IT BE HOT IN DJIBOUTI AT THE MOMENT] is unverified, or irrealis. The relation between a SR within the scope of *can*, *may*, *must* and *should* and the irrealis cognitive domain was shown to be inferred, albeit frequently, rather than procedurally encoded by these modals; with *will*, on the other hand, I shall argue that this relation is encoded.

By analogy to figures 7 and 11, *will*, on these assumptions, could be represented as relating a SR to the irrealis domain as in figure 17 below. (The irrealis domain is represented as a two-dimensional plane in view of my characterisation of it (§3.2.2) as incorporating varying degrees of likelihood of occurrence of events which may be either potential or hypothetical.)

Figure 17: the function of *will*



In relating a SR within its scope to the irrealis domain, *will* simply indicates that the event represented by the SR is unverified or potential. This is procedural information of the most general kind; in defending this characterisation of *will* I shall demonstrate that it does not encode any more specific information than this.

### 6.1.2 Will and temporal reference

We have already seen how *will* is unspecified with respect to the relation of SRs to assumptions in the utterance interpretation context, by comparing it with *must*. I shall also demonstrate that *will* is unspecified with respect to the temporal reference of SRs within its scope. *Will* does not encode future time reference; rather, the nature of the link between a SR within the scope of *will* and the temporal cognitive domain is either inferred or encoded by some temporal adverbial. Against this atemporal characterisation of *will* (although within a relevance theoretic procedural perspective)



- (6.9) A If Alison was born in April 1970, how old is she?  
 (6.7) B She'll be twenty six.  
 (6.10) A What, next April?  
 B No. I think she's twenty six now.

Maintaining a monosemous account, *will* is clearly not a future tense marker, although future time reference is often inferred from utterances of sentences containing *will*. The contribution made by tense in English is simply to distinguish between past and unspecified temporal reference. For example, the temporal reference of 'John will be working at home' (which is temporally unspecified) depends on the context in which it is used, in just the same way that unspecified tense marking alone underdetermines temporal reference:

- (6.11) a. There are no lectures today so John is working at home.  
 b. There are no lectures next week so John is working at home.  
 (6.12) a. There are no lectures today so John will be working at home.  
 b. There are no lectures next week so John will be working at home.

Because modals must be followed by an infinitive, past tense is indicated by the infinitive form of the auxiliary *have* following a modal auxiliary. This analysis is in line with McCawley's claim that "all occurrences of the auxiliary *have* are underlying past tenses" (McCawley 1971:99). So (6.13) and (6.14) locate the SR [JOHN BE WORKING AT HOME] at some time prior to the temporal reference point:

- (6.13) John was working at home.  
 (6.14) John will have been working at home.

Given that *will* encodes procedural information relating a SR within its scope to the irrealis cognitive domain rather than to the temporal cognitive domain (i.e. it specifies that the situation or event referred to is non-actual and unverified), why is it so often used to express future time events? Indeed this appears to be its prototypical use, as the classification of *will* as a 'future tense' in functional studies (e.g. Dahl 1985, Bybee *et al* 1994) and the use of stressed *will* to suggest future time reference in (6.15) suggest:

- (6.9) A If Alison was born in April 1970, how old is she?  
 (6.7) B She'll be twenty six.  
 (6.15) A What, she will be twenty six?  
 b No. I think she's twenty six now.

In §3.2.2 I stated that both the temporal domain and the irrealis domain are involved in the representation of future time reference, since a situation or event in the future is also logically irrealis, as figure 9, repeated below, illustrates:

Figure 9:

past / realis	present / realis	
past / irrealis	present / irrealis	future / irrealis

It is to be expected, therefore, that a marker of irrealis will be used to refer to futurate situations and events in a language with no future tense.

## **6.2 Will and shall**

### **6.2.0 Introduction**

We have seen that the semantics of *will* makes no reference to the relation of SRs to assumptions in the utterance interpretation context (in contrast to *can*, *may*, *must* and *should*) and does not specify temporal reference (so *will* is not a future tense marker). In this section, I shall demonstrate that *will* is ontologically unspecified with respect to the agent-event versus world-event distinction introduced in §3.3.3. In contrast, *shall* constrains the interpretation of SRs within its scope to agent-events. This procedural information subsumes a future time interpretation; in §6.2.2 I shall demonstrate that some of the differences in interpretation with *shall* with first person subjects (reflected in the traditional grammatical rule that future time should be expressed by *shall* with first person subjects and by *will* with others) reflects a partial generalisation of *shall* from being a marker of agent-events (subsuming future time reference) to a future time marker.

### **6.2.1 Agent-events and world-events**

In §3.3.3 I demonstrated that the use of the English imperfective (or progressive) aspect with future time reference is restricted to the expression of agent-events, that is, events brought about by "the motivated production of an activity with the intention of bringing a consequent state about" (Klinge, p.c.):

(3.73) Mary's having a party next week.



By the same token, English imperfective aspect with future reference cannot be used to refer to world-events, that is, events which come about simply by virtue of the way the world is, independent of any motivated human activity:

(3.46)           ? Mary's having a birthday next week.

In contrast, unmarked (simple) aspect with future time reference cannot be used to express agent-events in which the agent and the subject of the clause are coreferential:

(3.65)           Mary has a birthday next week.

(3.66)           ? Mary has a party next week.

Klinge (1993) links the agent-event versus world-event distinction to the expression of different types of modality (epistemic - concerning knowledge of and belief about events and situations; deontic - concerning permission and obligation; and dynamic - concerning volition and ability):

"Deontic [and also dynamic] modality is an unverified correspondence between a SITUATION REPRESENTATION of a WORLD SITUATION [a situation or event in the world] brought about by an AGENT-EVENT, epistemic modality is an unverified correspondence between a SITUATION REPRESENTATION and either an obtaining or [...] a previous WORLD SITUATION or a future WORLD-EVENT."

(Klinge 1993:346)

The distinction between deontic and dynamic modality depends on who motivates the agent-event; if the agent-event is motivated by the semantic subject of the clause a dynamic interpretation is recovered, but if the agent-event is motivated by someone other than the semantic subject a deontic interpretation is recovered.

The fact that *will* is used in the expression of all three categories of modality can be accounted for by proposing that utterances of sentences containing *will* are unspecified between agent-event interpretations and world-event interpretations. Inferential enrichment is required to determine what kind of modality (epistemic, deontic or dynamic), if any, an utterance of a sentence containing *will* is intended to convey, since *will* is semantically unspecified with respect to the agent-event versus world-event distinction.

In contrast to *will*, *shall* can never receive an epistemic interpretation. This leads Klinge (1993) to suggest that, whilst *will* can be used to refer to either an agent-event or a world-event (an 'unspecified-event'), *shall* indicates that a SR within its scope represents an agent-event rather than a world-event. So, in (6.16) Cinderella's going to

the Ball is dependent on the action either of the speaker (in which case (6.16) is interpreted as a promise) or of the addressee (in which case (6.16) is interpreted as a command) but can not be interpreted simply as a description of the way the world turns out to be:

(6.16) You shall go to the Ball, Cinderella!

Similarly in (6.17) the use of *shall* indicates that the situation described should be interpreted as an agent-event (with the additional implication that the agent in question is God rather than the meek):

(6.17) The meek shall inherit the Earth.

With *will* in place of *shall*, both (6.16) and (6.17) could (but need not) be simply predictions, that is, intended to be interpreted as world-events; this interpretation is not recoverable when *shall* is used. Thus (6.18) below would be infelicitous if *shall* replaced *will*:

(6.18) Will it work? You bet it will! (Mark Goodier, Radio 1, 21/3/1991)

This difference between *will* and *shall* manifests itself most obviously in interrogatives. Interrogatives with *shall* in the first person are often interpreted as offers or proposals with dynamic or deontic modality, whereas interrogatives in the first person with *will* are often requests for information with epistemic modality:

(6.19) Shall I visit John? (paraphrased as: 'Do you want me to visit John?' or 'Do you think it would be a good idea if I visited John?')

(6.20) Shall we visit John? (paraphrased as: 'Would you like us to visit John?' etc.)

(6.21) Will I/we visit John? (paraphrased as: 'Do you know if I/we will visit John?')

Similarly, interrogatives in the third person with *shall* are generally offers or suggestions, since the events referred to can only come about as the result of motivated human activity (agent-events), whereas interrogatives in the third person with *will* are generally requests for information. For example, Leech (1987:84) suggests (6.22) could be "spoken by a mother offering the services of her daughter to a neighbour", whereas (6.23) could not; (given an appropriate context, a volitional overtone can arise with *will*: "Is Gwen willing to do your shopping for you?"):

- (6.22)            Shall Gwen do your shopping for you?                            (Leech 1987:84)  
 (6.23)            Will Gwen do your shopping for you?                                (Klinge 1993:349)

Concerning interrogatives with *shall* in the second person, little can be said beyond the fact that they rarely (if ever) occur; Wekker's (1976) corpus of some 600,000 words (spoken and written) contained no examples of "Shall you...?" (Wekker 1976:58). Interrogatives with *will* in the second person are often requests or commands (referring to agent-events), but can be simply requests for information (possibly referring to world-events).

In summary, *shall* encodes information that the occurrence of a SR within its scope is due to an agent-event. A SR within the scope of *shall* represents a future situation or event, as *shall* is only ever used in future contexts in contemporary English; this, together with the observation that the agent-event versus world-event distinction only affects the distribution of English progressive and unspecified aspect with future temporal reference, suggests that the agent-event versus world-event distinction only operates where there is future time reference. The characterisation of *shall* as simply indicating that a SR within its scope represents an agent-event thereby incorporates reference to the future (abstract) cognitive domain.

### 6.2.2 *Shall*

*Shall* encodes information that a SR within its scope will be realised as an agent-event. Is this information procedural, or can it be characterised in terms of semantic retention, i.e., is the agent-event semantics of *shall* residual conceptual information formally encoded by its lexical source expression? In Old English, *scal* expressed obligation or necessity, even when expressing prediction, which as *shall* it increasingly did in Middle English (Oxford English Dictionary 1971:609.8). Bybee (1988:253) argues that as *will* developed as a marker of prediction in Modern English, it displaced *shall* from this function, leaving *shall* as a marker of obligation in all persons but as a marker of prediction only with first person subjects. This, she claims, shows that the development of *will* constricted the domain of application of *shall* but did not imbue it with any additional meaning, as only the source meaning of *shall* and meanings derivable from it contribute to the current semantic content of *shall*. However, as examples (6.16) and (6.17) illustrate, *shall* does not only express obligation, but can also be used to express a promise:

- (6.16)            You shall go to the ball, Cinderella!  
 (6.17)            The meek shall inherit the Earth.

This suggests that the information encoded by *shall* has changed from that encoded by its source; whether and to what extent this is due to the influence of contrast with *will* is unclear.

Still to be established is the status of the information encoded by *shall*: is it conceptual or procedural? The information encoded by *shall* that a SR within its scope represents an agent-event is recoverable whenever *shall* is used with second or third person subjects. With first person subjects an agent-event interpretation need not be recovered in declarative clauses, hence the acceptability of (6.24), but must be (at least in most dialects of contemporary British English) in interrogative clauses, hence the unacceptability of (6.25):

(6.24) I shall be twenty eight next year.

(6.25) ! Shall I be twenty eight next year?

There are two hypotheses to account for this anomaly. The first is that *shall* exhibits semantic retention, encoding conceptual information relating to agent-event status and procedural information relating a SR in its scope to the future abstract cognitive domain. However, this hypothesis fails to explain why semantic retention occurs in all declarative clauses apart from some with first person subjects.

The second hypothesis is that *shall* encodes procedural information to interpret a SR within its scope as representing an agent-event; subsumed by this information is an instruction to relate the SR to the future cognitive domain. The lack of an agent-event interpretation in some utterances with first person subjects in declarative clauses can be explained as evidence of generalisation, "the loss of specific features of meaning with the consequent expansion of appropriate contexts of use for a gram" (Bybee *et al* 1994:289), which as we saw in §5.1.2 is a mechanism which derives more general lexical meaning from specific lexical expressions, and more general grammatical meaning from grams, but does not itself derive grammatical meaning from lexical expressions. Generalisation can be described in terms of the loss of one feature of meaning, and this can easily be applied to *shall*:

*shall*

- (i) Agent-event: A SR within the scope of *shall* represents a future situation or event brought about by motivated human activity.
- (ii) Future: A SR within the scope of *shall* represents a future situation or event.

The reason this generalisation has been initiated with first person subjects in declarative clauses may be due to the fact that predictions in the first person often refer to agent-events (people have access to their own intentions that they do not have to others') with the result that first person declarative clauses with *shall* become generalised as first person predictions. Questions with *shall* in the first person refer to the intentions of others, as (6.19) and (6.20) illustrate (since there is no point in questioning one's own intentions, except rhetorically) and therefore have not been generalised as first person predictions.

This suggests that *shall* encodes procedural information that a SR within its scope represents an agent-event, and that this has undergone the first stage of generalisation towards expressing future time reference independent of any agent-event interpretation. However, there are reasons to expect that this process of generalisation will not be completed. In Middle English, according to the OED (1971:609.8) "the future was expressed by either *shall* or *will*, the former being much more common." *Shall* has since been largely displaced by *will* as a marker of prediction in all persons, particularly in North American varieties of English; Poplack & Tagliamonte (1995) found only two tokens of *shall* out of a total of 2,651 futurate clauses in their study of the forms used for future temporal reference in three Canadian speech communities.<sup>63</sup> Decreased frequency of an expression is accompanied by a restriction rather than an expansion of appropriate contexts of use, suggesting that the generalisation of *shall* which appears to have been initiated has been curtailed by the development of *will*, and possibly also by that of the more recent *be going to* construction.

### 6.3 Will and be going to: scopal differences

#### 6.3.0 Introduction

In this and the following section I shall compare the two currently most common future expressions in English, *will* and *be going to*. These exhibit a number of differences both in distribution and in interpretation. I shall begin by arguing (against Klinge 1993) that *be going to* is not disqualified from classification as an operator on the grounds that its scope differs from that of *will* and shall then demonstrate that differences between *will* and *be going to* in respect to their interaction with past-time temporal adverbials can be accounted for in terms of their relative scope, rather than (as Klinge 1993 claims) semantically. I shall conclude by demonstrating that

---

<sup>63</sup> The communities studied were North Preston African-Americans, Guysborough African-Americans and Guysborough Whites; the two instances of *shall* came from the Guysborough African-American community (total of futurate forms 1,264).

differences between *will* and *be going to* in respect to their co-occurrence with *already*, which Haegeman (1989) attributes to pragmatic factors, can also be accounted for in terms of relative scope. Both discussions shed light on the semantics of *be going to*, which will be taken up in §6.4.

Both Klinge (1993) which provides an analysis of the English modal auxiliaries, briefly mentioning *be going to*, and Haegeman (1989) which provides a pragmatic account of *will* and *be going to*, assume a relevance theoretic framework. However, Klinge (1993) and Haegeman (1989) differ concerning the status of *be going to* with respect to the conceptual/procedural distinction. Haegeman analyses both *be going to* and *will* as exponents of procedural encoding; she writes:

"*Be going to* ... imposes a constraint on the processing of the proposition with which it is associated. It signals that this proposition is relevant in a context including at least some present tense propositions, or, in other words, it guarantees a contextual effect if the utterance is processed against a present context. *Will*, on the other hand, signals that the hearer should extend the immediately accessible (present) context for the processing of the proposition and should process the utterance against future propositions."

(Haegeman 1989:305)

Klinge (1993) also holds that *will* is procedural but argues that it, along with all the modals, encodes "potentiality" (1993:323) rather than futurity as the constraint imposed on processing (§4.2.1 discusses Klinge's analyses of *may* and *must*). In other words, the proposition with which *will* is associated is to be processed as potential rather than actual. Since future contexts are logically also potential, Klinge argues (as I did in §6.1.2 against Haegeman 1989) that although *will* is often processed against a future context, this is not what it encodes; in other words, futurity is not part of the semantics of *will*, even though *will* can be used in the expression of future time propositions. As for *be going to*, Klinge (1993:346) assumes, for reasons to be discussed below, that this encodes conceptual information. The differences between the characterisations of *will* and *be going to* in Haegeman (1989) and Klinge (1993) are summarised in figure 18:

Figure 18:

	Procedural	Conceptual
<b>Klinge (1993)</b>	<i>will</i> 'potentiality'	<i>be going to</i>
<b>Haegeman (1989)</b>	<i>will</i> 'future context' <i>be going to</i> 'present context'	

I shall first discuss and reject Klinge's claim that *be going to* does not encode procedural information, and then demonstrate that one of Haegeman's arguments purporting to demonstrate that *be going to* encodes procedural information instructing addressees to process utterances against present time contexts can in fact be accounted for as a scopal difference relating to the relative scope of *will* and *be going to*.

### 6.3.1 *Be going to* as an operator

The *be going to* construction exhibits many of the features of grammatical markers discussed in §5.1.1. It is semantically bleached, not being restricted, for example, to co-occurrence with subjects capable of movement towards a goal:

(6.26) It's going to rain.

It has a phonetically reduced allomorph, *gonna*, and exhibits a high degree of syntactic rigidity, in that it only takes bare non-finite verbal complements, in contrast to the non-grammaticalized present progressive form of *go* with locative *to* in (6.27a):

- (6.27) a. I'm going to town to see a film.  
b. \* I'm gonna town to see a film.  
c. I'm gonna see a film in town.

Being grammaticized, *be going to* can be characterised as an operator, that is, as an exponent of procedural encoding at phase two with scope over SRs. However, Klinge (1993) characterises the *be going to* construction as an exponent of conceptual rather than procedural information. Klinge's comments on *be going to* are brief and are repeated below in full:

"*Be going to* does not seem to occur where the WORLD SITUATION is in past time relative to the time of utterance:

(6.28) \* They are going to have arrived two hours ago.

Apparently, the explanation is that *be going to*, like *be willing to* is part of the SITUATION REPRESENTATION, as suggested by the following example from Leech (1987:57):

(6.29) Call me at lunch time on Monday - I'll be going to speak to the boss about it that afternoon.

with the structure:

(6.30) WILL [I BE GOING TO SPEAK TO THE BOSS]"

(Klinge 1993:346)





We can therefore talk of the unique PC encoded by an utterance but not of the unique SR (unless only one operator is encoded in the utterance). The PC consists solely of linguistically encoded conceptual information, whereas a SR within the scope of a given operator, O, consists of the conceptual information encoded by the PC inferentially enriched in line with the procedural information encoded by any other operators with narrower scope than O.

Given this framework, when *be going to* falls within the scope of *will*, it is part of the SR over which *will* has scope (as Klinge 1993 claims), but this does not prevent *be going to* itself from functioning as an operator (and therefore as an exponent of procedural encoding) with scope over a further SR:

(6.30) WILL [I BE GOING TO SPEAK TO THE BOSS]

(6.30') WILL [BE GOING TO [I SPEAK TO THE BOSS]]

Klinge's other argument for treating *be going to* as part of a SR was that whilst *will* can occur when the temporal reference of the SR within its scope is prior to the time of utterance, *be going to* can not:

(6.28) \* They are going to have arrived two hours ago.

(6.31) They will have arrived two hours ago.

This observation does not in itself demonstrate that *be going to* is not an operator, or that it cannot encode procedural information; it simply demonstrates that the distribution of *will* and *be going to* are not co-extensive. In the remainder of this subsection I shall attempt to account for the acceptability of (6.31) and the unacceptability of (6.28) within the framework outlined above.

Temporal adverbials such as *two hours ago* behave similarly to the manner adverbials (*seriously, frankly, confidentially*) discussed in §2.3.2, in that they sometimes exhibit scopal ambiguity. In (6.33) below, the temporal adverbial *two hours ago* can modify either the saying or the being here:

(6.33) She said she would be here two hours ago.

When analysing the logical forms of examples such as (6.28) and (6.31) I shall adopt a 'minimal attachment' approach and incorporate the temporal adverbial within the nearest SR, i.e. the last or innermost SR. If this yields a contradictory result, I shall incorporate the temporal adverbial within the next SR and so on until no contradiction

arises. If an acceptable sentence yields a contradictory result with the temporal adverbial modifying all SRs, or if an unacceptable sentence does not result in a contradictory logical form, my analysis will have been proved untenable.

In §3.3 I characterised tense markers as operators, constraining the manipulation of SRs with respect to the temporal cognitive domain; I shall argue here that it is the procedural semantics of *will* and *be going to*, and the scope of each relative to the tense operator which accounts for the acceptability of (6.31) and the unacceptability of (6.28). In (6.31), the past<sup>64</sup> tense operator encoded in *have arrived* is within the scope of *will*, rather than having scope over it; the temporal adverbial has been minimally attached to the innermost SR (SR1):

- (6.31) a. They will have arrived two hours ago.  
b. WILL [2 PAST [1 THEY ARRIVE TWO HOURS AGO]]

The PAST operator in (6.31) locates the event [THEY ARRIVE TWO HOURS AGO] in the past (it is therefore functionally redundant, but its omission would be incompatible with the temporal adverbial phrase), and WILL encodes information to interpret the SR within its scope as potential (not explicitly future) rather than verified.

Turning to (6.28) we see that, unlike *will*, *be going to* is within the scope of a tense operator encoded in the finite verb form *are* and has scope over a further SR or SRs:

- (6.28) a. \* They are going to have arrived two hours ago.  
b. NON-PAST [3 BE GOING TO [2 PAST [1 THEY ARRIVE TWO HOURS AGO]]]

Let us assume that *be going to* functions as some kind of future marker, i.e. it encodes information to process the SR within its scope as occurring subsequent to the time set up by the tense of the finite form of *be* (in §6.4 I shall defend this characterisation of the procedural information encoded by *be going to*). Given this assumption, the SR of which *be going to* is part (SR3) must be temporally prior to SR2; that is, SR3 < SR2.

---

<sup>64</sup> Klinge (1993:345) refers to the tense in (6.31a) as the perfect and renders my (6.31b) as his (75):

(75) WILL [PREVIOUS [THEY ARRIVE TWO HOURS AGO]]

I have used PAST rather than PREVIOUS because in sentences such as (6.31a) the past/perfect distinction is lost due to the grammatical constraint to follow a modal auxiliary with a non-tensed form. Also, if the modal auxiliary were omitted from (6.31a) the resulting form would be in the past tense (9) rather than the perfect (10):

- (9) They arrived two hours ago.  
(10) \* They have arrived two hours ago.

Now, in (6.28) the outermost (NON-PAST) operator restricts the temporal reference of SR3 to a time coreferential with or subsequent to the time of utterance ( $t_0$ ):  $SR3 \geq t_0$ . Now, SR2 is located at a time two hours prior to  $t_0$ :  $t_0 > SR2$ , which entails that SR3 is subsequent to SR2:  $SR3 > SR2$ . This gives  $SR3 < SR2$  and  $SR3 \geq t_0 > SR2$ . Since SR3 cannot be both prior to and subsequent to SR2, (6.28) is contradictory. This holds whether the temporal adverbial [TWO HOURS AGO] is incorporated within SR1, SR2 or SR3.

In contrast, if the tense operator with scope over *be going to* is PAST, the resulting sentence is acceptable:

- (6.34) a. They were going to have arrived two hours ago.  
 b. PAST [<sub>3</sub> BE GOING TO [<sub>2</sub> PAST [<sub>1</sub> THEY ARRIVE TWO HOURS AGO]]]

The following temporal relations amongst the SRs are encoded in (6.34):

Because of the procedural content of *be going to*, SR3 must be temporally prior to SR2, hence  $SR3 < SR2$ .

SR3 is within the scope of a deictic PAST tense operator, hence  $SR3 < t_0$ .

Because of the temporal adverbial in SR1 and the PAST operator with scope over it, SR1 (and therefore SR2 also) is prior to  $t_0$ , hence  $SR2 < t_0$ .

These relations are compatible as follows:  $SR3 < SR2 < t_0$ , hence (6.32) ought to be acceptable, which it is.

The most natural interpretation of (6.34) would be that the subject of the clause did not in fact arrive two hours ago, whereas (6.31), with *will*, suggests that they did arrive two hours ago (although this is not known for certain). Relevance theory accounts for this through the principle of relevance: "Every act of ostensive communication communicates the presumption of its own optimal relevance" (S&W 1986:158), where optimal relevance is defined as adequate contextual effects for minimum processing effort. The most economical way of stating that they did not arrive two hours ago is to say just that; however, the most economical way of stating that they were expected to arrive, had planned to arrive, etc. two hours ago, but didn't, might be to utter (6.34a).<sup>65</sup>

---

<sup>65</sup> There is an alternative reading of (6.34a) for some people, in which the potential for, or expectation of, the SR [THEY HAVE ARRIVED], rather than the SR itself, is located two hours prior to the utterance. In this reading, the temporal adverbial [TWO HOURS AGO] modifies SR3. This is represented in (11) below with an accompanying paraphrase:

- (11) PAST [<sub>3</sub> BE GOING TO TWO HOURS AGO [<sub>2</sub> PAST [<sub>1</sub> THEY ARRIVE]]]  
 Two hours ago, they were going to have arrived.

In summary, I first demonstrated that although *will* has wider scope than *be going to* this does not bar *be going to* from functioning as an operator in its own right; *be going to* is an operator with narrower scope than *will* and can contribute to a SR within the scope of *will*. I then demonstrated that the contrast between the unacceptability of (6.28) with *be going to* and the acceptability of (6.31) with *will*, which Klinge (1993) attributed to semantic differences between *be going to* and *will* with respect to the conceptual/procedural distinction, could be accounted for in terms of relative scope. In discussing these examples, I assumed that *be going to* encodes procedural information to process the SR immediately within its scope as subsequent to the SR to which *be going to* itself contributes. In contrast, *will* encodes procedural information concerning atemporal potentiality, hence the acceptability of (6.5) with *will* but not of (6.35) with *be going to*:

(6.5)           Bleach will kill germs.

(6.35)          ? Bleach is going to kill germs.

### 6.3.2 Co-occurrence restrictions with *already*

To conclude this section, I shall demonstrate that scopal, as opposed to pragmatic, differences between *will* and *be going to* contribute to the different co-occurrence patterns of *will* and *be going to* with the temporal adverbial *already*. These co-occurrence patterns also shed more light on the semantics of *be going to*.

Haegeman (1989) claims that *will* encodes procedural information to process an associated utterance against a future context, whilst *be going to* encodes procedural information to process an associated utterance against a context including at least some present tense assumptions. In §6.1.2 I demonstrated the untenability of Haegeman's characterisation of *will*, and in this section I shall demonstrate that one of the arguments purporting to support the above characterisation of *be going to* does not in fact do so.

According to Haegeman (1989) the differences between *will* and *be going to* are pragmatic; they impose constraints on the construction of utterance interpretation contexts. Haegeman (1989:296) proposes that the contrast between *will* and *be going to* might be clarified by looking at their co-occurrence with *already*, which, she claims, also "imposes 'pragmatic' constraints on processing." Specifically, Haegeman (1989:296) claims:

1) that *already* relates events or states to other events in the immediate (i.e. present) context rather than to a point in time (at least where present time is concerned).

2) that *already* therefore occurs with *be going to* rather than with *will*, since *be going to* guarantees adequate contextual effects from a proposition if it is related to a present context.

Now, in §1.3.3 I illustrated how the term 'pragmatic' can be used to mean either inferred (as opposed to encoded), or non-truth-conditional, and how these definitions need not be coextensive. Since the differences discussed in Haegeman (1989) are differences between two linguistically encoded expressions, 'pragmatic' can only be being used to mean non-truth-conditional. However, *already* is not 'pragmatic' in the sense of non-truth-conditional; *already* indicates that a situation holds not only at the temporal reference point (for example, five p.m.) but also prior to then:

- (6.36) a. I was here at five p.m. and not before five p.m.  
b. \* I was already here at five p.m. and not before five p.m.<sup>66</sup>

This information is truth-conditional, that is, it contributes to the proposition expressed. (Example (6.36) also illustrates that *already* can relate situations to points in past time (five p.m.) and does not have to relate them to other events when the temporal reference is past.)

Against Haegeman's second claim, that *already* occurs with *be going to* rather than with *will*, consider the following examples:

- (6.37) He's already going to visit John.  
(6.38) \* He'll already visit John.  
(6.39) He'll already be visiting John.

(6.38) only differs from (6.39) in that the SR [HE VISIT JOHN] is within the scope of a PROGRESSIVE operator in (6.39). (6.38) is unacceptable because *already* can only co-occur with a SR with progressive or perfect aspect since *already* imposes constraints on the temporal reference of states, or situations, rather than events - ongoing with progressive aspect or resultant with perfect aspect (cf. Michaelis 1996). In the logical forms below, ALREADY has been represented in the same way as the temporal adverbial TWO HOURS AGO which is also truth-conditional. However, bearing in mind that *already* is restricted to co-occurrence with states rather than events, it has been incorporated into the innermost SR having PROGRESSIVE aspect (except in the unacceptable (6.38) where there is only one SR):

---

<sup>66</sup> I am grateful to Vladimir Zegarac for these examples.

- (6.37) a. He's already going to visit John.  
 b. NON-PAST [HE BE GOING TO VISIT JOHN ALREADY]  
 c. NON-PAST [<sub>2</sub> BE GOING TO ALREADY [<sub>1</sub> HE VISIT JOHN]]
- (6.38) a. \* He'll already visit John.  
 b. WILL [<sub>1</sub> HE VISIT JOHN ALREADY]
- (6.39) a. He'll already be visiting John.  
 b. WILL [HE BE VISITING JOHN ALREADY]  
 c. WILL [<sub>2</sub> PROGRESSIVE ALREADY [<sub>1</sub> HE VISIT JOHN]]

In (6.37), ALREADY can not be constraining the temporal reference of the event represented by SR1, since this is not a state; it must therefore be interacting directly with SR2, containing BE GOING TO, which must therefore be a representation of a state rather than an event. ALREADY indicates that the situation described by SR2 holds at and prior to  $t_0$ , and the procedural semantics of BE GOING TO locates SR2 prior to SR1. (6.37) is acceptable because these temporal relations (represented as follows) are compatible:  $SR2 \leq t_0 < SR1$ .

In (6.39), WILL has scope over the SR containing ALREADY, indicating that this represents a potential situation (either in the present or in the future). Because of the semantics of ALREADY, SR2 holds at and prior to the temporal reference set up, for example, by an accompanying temporal adverbial phrase such as “when you try to phone him”.

To summarise, the different co-occurrence patterns of *will* and *be going to* with *already* do not demonstrate, as Haegeman (1989) claims they do, that *will* and *be going to* are only pragmatically (i.e. non-truth-conditionally) distinct. The logical forms above show that ALREADY combines with SRs representing situations or states. WILL does not contribute to a SR as modal auxiliaries in English are the outermost operators, and so ALREADY falls within the scope of WILL. BE GOING TO, on the other hand, does contribute to SRs containing ALREADY; this demonstrates that *be going to*, as well as encoding procedural information relating to future time reference (as the analysis of (6.28) and (6.31) illustrated), can also represent an ongoing state. This is not surprising given that the *be going to* construction is derived from the progressive form of the verb *go* plus the allative marker *to*. In §6.4 I shall examine the semantics of *be going to* in terms of the conceptual/procedural distinction.

## 6.4 Will and be going to: semantic differences

### 6.4.0 Introduction

In this section I shall begin by providing a justification for the characterisation of *be going to*, assumed in §6.3, as encoding procedural information relating SRs to the future (abstract) cognitive domain. I shall then discuss differences in distribution and interpretation between *will* and *be going to* which cannot be accounted for purely in terms of differences between their relative scope, and suggest that these differences arise when *be going to* exhibits semantic retention. I shall conclude by specifying the conceptual semantic content encoded by *be going to*.

### 6.4.1 Procedural information encoded by *be going to*: a justification

The previous section demonstrated that *be going to* can be analysed as an operator, that is, as an exponent of procedural encoding at phase two of utterance interpretation with scope over SRs. I assumed that the procedural information encoded by *be going to* concerned futurity, or (strictly) ulterior time reference, that is, future time relative to some temporal reference point (R) which need not be the time of utterance ( $t_0$ ). In other words, a SR within the scope of *be going to* (SR1) must be related temporally to a time subsequent to that of the SR containing *be going to* (SR2); that is:

OPERATOR [<sub>SR2</sub> BE GOING TO [<sub>SR1</sub> ...]] where  $SR2 < SR1$ .

The temporal reference (R) of SR2 may be indicated by a temporal adverbial, such as 'two hours ago', or may be inferred, the inference being constrained by a temporal or modal operator with scope over SR2, such as the past tense form of *be*.

This characterisation of *be going to* as a future time marker arises from the following considerations. First, it has been noted (Fries 1927, Ultan 1978, Bybee *et al* 1991:18, Bybee *et al* 1994:244) that, cross-linguistically, future grams typically develop from movement verbs, from lexical constructions expressing obligation, desire, and ability, and, less commonly, from temporal adverbs. Whilst *shall* has evolved from a lexical source encoding obligation, the *be going to* construction has evolved from a movement verb, *go*. Secondly, Bybee *et al* (1994:11) observe that future grams typically develop from constructions with imperfective rather than perfective or past marking (encoded in the *be going to* construction by *be +ing*) and from constructions incorporating an allative component (here encoded by the preposition *to*).<sup>67</sup> Since

---

<sup>67</sup> The relation between allatives and future time reference can be illustrated by comparing allative *to* with the gerundive *-ing*:

(12) a. You must remember to lock the door.

movement verbs are also a common source of pasts and progressives, Bybee *et al.* (1991:30) suggest that movement alone is not sufficient to give rise to future grams; imperfectivity and allativity are also required. It is therefore the combination of movement verbs plus imperfective marking and an allative component, rather than movement verbs alone, that gives rise to future grams; hence, the semantic content of the entire pre-grammaticized *be going to* construction contributes to the procedural semantics of the resulting operator.

Given that the *be going to* construction exhibits the formal features of grammaticization and is derived from the imperfective form of a movement verb with an allative complement, we would expect it to encode procedural information relating to future time reference. In support of this characterisation is the fact that in §6.3.1 the analysis of *be going to* as an operator encoding future time reference allowed us to account for the unacceptability of (6.28) in contrast to the acceptable (6.31):

(6.28) \* They are going to have arrived two hours ago.

(6.31) They will have arrived two hours ago.

Similarly, we would expect that if *be going to* encodes procedural information constraining the interpretation of SRs with respect to future temporal reference, it might be interchangeable with *will* with future time interpretation; and this is the case:

(6.40) a. Hurry up! We're going to be late!

b. Hurry up! We'll be late!

(6.41) a. The world's oil supply is going to run out one day.

b. The world's oil supply will run out one day.

Whilst *be going to* and *will* can be used interchangeably in the above examples, some peripheral contrasts in interpretation may be observed; that is, various 'overtones' associated with the use of *will* or *be going to* can be recovered, if licensed by the utterance interpretation context. For example, McIntosh (1966:307-309) suggests that examples such as (6.40a) stress 'present orientation', for example that the addressee is dawdling at the moment, whereas (6.40b) stresses the contingent nature of the future event, i.e. something can still be done to avoid being late; examples like (6.41a)

---

b. You must remember locking the door.

In (12a) the event YOU LOCK THE DOOR is envisaged subsequent to the time of utterance, whilst in (12b) this event is envisaged prior to the time of utterance. I am grateful to David Adger for bringing this distinction to my attention.



would, according to McIntosh (1966) emphasise the inevitability of the world's oil supply running out, whereas (6.41b) would emphasise the fact that this event is still a long way off. However, speakers are not always interested in conveying these additional overtones, and in such cases it would not be worth the addressee recovering them since optimal relevance would be achieved simply by recovering a proposition along the lines of "at some future time, x".

The possible differences in interpretation between the minimal pairs with *be going to* and *will* in examples (6.40) and (6.41), are very subtle, and some speakers of English fail to make these distinctions at all. Below, I shall look at further minimal pairs in which differences in interpretation and distribution between *be going to* and *will* with future time reference are clearer. I shall continue to describe these differences as arising from 'overtones' associated with *will* or *be going to*, since 'overtone' is a theoretically neutral term. In the following discussion, I shall determine whether these overtones, which I shall argue are the result of semantic retention in the case of *be going to*, are truth-conditional or not, and whether they are explicatures or implicatures.

#### 6.4.2 A descriptive characterisation of *be going to*

In this section, I shall illustrate some of the overtones associated with utterances of sentences containing *be going to* in contrast to *will*. The use of *be going to* in various contexts can suggest prior intention, inevitability or imminence; the following minimal pairs illustrate these three overtones associated with *be going to*.

##### *Prior intention*<sup>68</sup>

If an affirmative response to a request contains *will*, intention subsequent to the request is suggested; if the response contains *be going to* intention prior to the request is suggested, as the following minimal pair demonstrates:

- Can somebody visit John tomorrow?
- (6.42) a. I'm going to visit him.  
b. I'll visit him.

Intention to visit John originating prior to the request is indicated in (6.42a), that is, the speaker of (6.42a) was already intending to visit John at the time of the request; (6.42b), on the other hand, indicates intention to visit John originating subsequent to

---

<sup>68</sup> Also termed "intention with forethought" (Coates 1983:200), "present decision or intention" (Palmer 1979:122) and "future culmination of present intention" (Leech 1987:59).

the request, that is, the speaker of (6.42b) had not been intending to visit John but decides to on being asked.

The difference between intention prior to a request (associated with *be going to*) and intention subsequent to a request (associated with *will*) also affects the interpretation of interrogatives. Interrogatives in the second person with *will* are often requests (6.43) or orders (6.44) (see §5.3.2), but such interrogatives with *be going to* are more often questions about the addressee's prior intentions (6.45):

- (6.43) Will you do the washing up for me, please?  
(6.44) Will you please stop making that awful noise!  
(6.45) Are you going to do the washing up, or am I?

The final sets of minimal pairs illustrate the effect of the overtone of prior intention on the acceptability of *be going to* in certain contexts. In the following situations, only (6.46) and (6.48) are acceptable as it is impossible to intend to answer a telephone before it rings, or to intend definitely to ask for help before help is needed:

- Telephone: 'Rrring, Rrring' (immediately):  
(6.46) I'll get it!  
(6.47) ? I'm going to get it!  
(6.48) I promise I'll ask you if I need help.  
(6.49) ? I promise I'm going to ask you if I need help.

### *Inevitability*<sup>69</sup>

Inevitability is suggested in (6.50a) with *be going to* in contrast to contingency in (6.50b) with *will*:

- (6.50) a. Don't go near that parcel! It's going to explode!  
b. Don't go near that parcel! It will explode!

(6.50a) suggests that the parcel is going to explode whether anyone goes near it or not, whereas (6.50b) suggests that the parcel will explode only if someone goes near it.

---

<sup>69</sup> Inevitability is often connected in the literature with the notion of "future culmination of present cause" (Lecch 1987:59). Palmer (1979:123) explicitly links the two when he says: "Often there is a sense of inevitability - the train of events is already in motion." This is not a necessary connection however; *be going to* can be used to stress the inevitability of future events without explicit reference to their present-orientation, as in (6.41a) above: 'The world's oil supply is going to run out one day'.

Because of the contingent nature of many *will* constructions, *will* rather than *be going to* is generally found in the apodosis of conditional sentences.<sup>70</sup> *Be going to* can be used in the protasis of conditional clauses where, as Palmer (1979:124) notes, it contrasts with the unmarked form:

(6.51) If she stays more than one night, she can have my room.

(6.52) If she's going to stay more than one night, she can have my room.

In (6.51) her having the room is dependent on and subsequent to her staying more than one night, which is contingent; in (6.52) her staying more than one night is inevitable, or at least confidently envisaged. *Be going to* can also be used in the protasis of a conditional clause to indicate the continuation of events, as well as the planned nature of events as in (6.52); (6.53) indicates that running around is already in progress and likely to continue:

(6.53) If you're going to run around like that, do it outside!

#### *Imminence*<sup>71</sup>

Generally, if events are not imminent, and especially if they are envisaged subsequent to another future event, e.g. "After they build new houses...", *will* is used:

(6.54) After they build new houses here, they'll have to widen the roads and put up gas stations, and pretty soon the whole area will just be a big strip.

(Watterson 1988 *Something Under the Bed is Drooling* p.99)

But *will*, being modal, cannot be used for 'future in the future' as no modal verb can follow another verb, so 'future in the future' takes the form: "...will be going to...".

In the examples below, imminence is suggested in (6.55a) but not necessarily in (6.55b). Whereas (6.55b) is what Binnick (1972:7) terms 'elliptical' (that is, it is incomplete as it stands) unless qualified in some way ("...if I eat any more ice-cream", "...after an hour on a coach"), (6.55a) is not 'elliptical' because being sick is envisaged as inevitable and imminent:

---

<sup>70</sup> This observation is corroborated by data from all three corpora in Poplack & Tagliamonte (1995) which show *will* occurring in over 60% of main clauses of conditional sentences.

<sup>71</sup> Also referred to as "immediate future with no current activity" (Palmer 1979:122) and "a time immediately subsequent to the moment of speaking" (Coates 1983:204).

- (6.55) a. I'm going to be sick!  
b. I'll be sick!

The idea of imminence is often associated, as in (6.55a), with the overtone of inevitability (indeed Leech (1987:60) subsumes imminence and inevitability under 'current activity leading to a future event'), and also with the overtone of prior intention; for example, Palmer's (1979:122) example of "immediate future with no current activity":

- (6.56) What are you going to play us?

conveys the overtone of prior intention, paraphrased as, 'What have you planned to play us?'. Nonetheless, I shall include imminence as a separate category because, as McIntosh (1966:306) states:

"Distinctions between present- and future-orientation are often of no great importance;... But more generally critical is the choice of G-constructions [forms of *be going to*], other things being equal, for the expression of something imminent: She's going to faint."

There is, however, considerable variation amongst speakers over the accessibility of the overtone of imminence associated with *be going to*. Poplack & Tagliamonte (1995) found that *be going to* was favoured for immediate future reference (defined as events, processes and states understood - on the basis of contextual information such as adverbial specification - to have occurred between ten seconds and one hour after the utterance) in one of their three corpora ("Guysborough Whites"). In this corpus, *be going to* occurred 44% of the time in imminent contexts but only 18% of the time elsewhere, whereas the other two corpora demonstrated no preference for *be going to* in imminent contexts (in line with Myhill 1994). Whilst these findings do not demonstrate that *be going to* encodes imminence for Guysborough Whites, they do indicate a positive functional correlation between imminence and the use of *be going to* in this speech community. This provides a counterexample to the findings of Bybee *et al* (1994) - for whom grammatical categories are functionally determined<sup>72</sup> - when they write:

"The most common sources of immediate futures in our database are verbs meaning 'come to'; 'go to' sources do not yield immediate futures in our data."

(Bybee *et al* 1994:280)

---

<sup>72</sup> For example, *will* is classified as a future gram because of its high correlation with future contexts, although, as I demonstrated in §6.1.2, *will* is not a future tense marker.

### Summary

Three overtones associated with the use of *be going to* have been illustrated: prior intention (6.42a), inevitability (6.50a), and imminence (6.55a).

- Can somebody visit John tomorrow?
- (6.42) a. I'm going to visit him.
- (6.50) a. Don't go near that parcel! It's going to explode!
- (6.55) a. I'm going to be sick!

The theoretical status of these overtones has yet to be established; in the following section I shall determine whether these overtones are truth-conditional or not, and whether they are explicatures or implicatures.

#### 6.4.3 The pragmatics of *be going to*

In this section I shall determine the theoretical status of the different overtones typically associated with utterances of sentences containing *be going to* (prior as opposed to subsequent intention, inevitability as opposed to contingency, and imminence). In terms of their logical semantic/pragmatic status, are the overtones part of 'what is literally said', i.e. do they contribute to the truth-conditions of the associated utterance, or are they non-truth-conditional? And from the linguistic semantic/pragmatic point of view, in which 'semantic' is defined as encoded and 'pragmatic' as inferred (this is the usage adopted here), are the overtones part of the explicit content of utterances, or are they implicatures? I shall address these questions in turn.

First, are the overtones associated with certain utterances containing *be going to* truth-conditional, that is, do they contribute to the proposition expressed? Consider the following examples:

- Can somebody visit John tomorrow?
- (6.57) ? I'm going to visit him but I wasn't intending to.
- (6.58) Don't go near that parcel! It's going to explode!  
? But if you keep well clear it won't explode.
- (6.55) a. I'm going to be sick!
- (6.59) If he's going to be sick, we'd better stop the car **right now**.

In (6.57) and (6.58), cancellation of the overtones of prior intention in (6.57) and inevitability in (6.58), leads to a sense of contradiction, suggesting that prior intention

and inevitability are part of the propositions expressed by utterances of (6.57) and (6.58). In (6.59), the overtone of imminence can enter into a relationship of entailment, suggesting that this also is part of the proposition expressed rather than a non-truth-conditional implicature. The conclusion 'we'd better stop the car right now' only follows from 'he's going to be sick' if his being sick is an imminent event rather than an event expected at some unspecified time in the future.

The fact that these overtones contribute to the propositions expressed in the above examples, does not entail that they are encoded. As already noted, relevance theory recognises pragmatic aspects of what is said (truth-conditional content). The second question to be answered, therefore, is whether the overtones associated with certain utterances containing *be going to* are encoded by *be going to* or inferred.

If they are directly encoded, then *be going to* must be polysemous, given that not all of the overtones are recovered in every utterance, as the examples above illustrate. However, there are significant problems associated with a polysemous account. To begin with, it is not clear what the basic meanings of *be going to* would be. For example, Coates (1983:198-204) distinguishes between root and epistemic "meanings", whilst Leech (1971:54) distinguishes "between two meanings [of *be going to*], the FUTURE CULMINATION OF PRESENT INTENTION and the FUTURE CULMINATION OF PRESENT CAUSE." Even if two meanings could be distinguished, there is still the problem of accounting for utterances in which none of the overtones are evident (suggesting a further meaning of *be going to*, that of pure future).

There is also the problem of how to account for utterances like (6.60) which can be interpreted in more than one way, as evidenced by the possible responses (6.61) to (6.63): The interlocutors A and B are expecting visitors for a meal; B is standing in the kitchen filling a kettle; A enters:

- (6.60) A: Are you going to make a pot of tea?  
(6.61) B: No, I'm going to boil some water for the potatoes.  
(6.62) B: No, I'll wait / I'm going to wait until the guests have arrived.  
(6.63) B: Yes, when the guests arrive.

In (6.61) B has interpreted A's question as referring primarily to the current activity (filling the kettle), and the response is intended to be interpreted likewise, as a reference to the current activity; B's intention to make a pot of tea is neither denied nor affirmed, despite the answer, 'No...'. In (6.62) B has interpreted A's question as referring to the imminence or otherwise of making a pot of tea so the response 'No...'

denies that a pot of tea is on the point of being made, although the remainder of (6.62) affirms that B intends making one; nothing is said concerning the current activity. In (6.63) B has interpreted A's question as referring to B's intentions; the current activity is incidental and nothing is said concerning it. If (6.60) were genuinely ambiguous the speaker (A) would have intended to communicate one specific proposition which the addressee (B) could be expected to recover, and only one of the responses above would be felicitous. As it is, any of the responses (6.61) to (6.63) is felicitous given an appropriate interpretation of (6.60); this suggests that (6.60) encodes a single logical form which the addressee can enrich in such a way as to license an optimally relevant response.

If (6.60) encodes a single logical form, *be going to* must be monosemous and therefore the specific interpretations of (6.60) - reflected in the responses (6.61) to (6.63) - arise from (6.60) being inferentially enriched in different ways. The overtones of prior intention, inevitability and imminence cannot, therefore, be directly encoded by *be going to*. But neither can the overtones of prior intention and inevitability be purely inferred, given that they arise with utterances containing *be going to* but not *will*:

Can somebody visit John tomorrow?

- (6.42) a. I'm going to visit him.  
b. I'll visit him.

- (6.50) a. Don't go near that parcel! It's going to explode!  
b. Don't go near that parcel! It will explode!

Imminence, however, can arise as an overtone of an utterance containing *will*. Imminence is conveyed in (6.46); *will* is felicitous but *be going to* in (6.47) is not, because the use of *be going to* could suggest an intention to answer the phone prior to it ringing, which is impossible:

Telephone: 'Rrring, Rrring' (immediately):

- (6.46) I'll get it.  
(6.47) ?I'm going to get it.

Also, the overtone of imminence can be easily cancelled even where prior intention or inevitability cannot; in Palmer's (1979:122) example of "immediate future with no current activity", (6.56), the assumption that the addressee has already decided what to play can be recovered whether (6.56) is uttered a month prior to a concert or when

the addressee is seated at the piano and about to begin to play. The overtone of imminence is only recoverable given the latter context, and even then can be cancelled, as (6.64) illustrates:

(6.56) What are you going to play us?

(6.64) Some Chopin, but first I'd like to say a few words...

Similarly, inevitability is suggested whether (6.65) below is uttered by someone reading a long-term weather forecast or by someone who can already see clouds gathering; imminence is only suggested in the latter scenario:

(6.65) It's going to rain.

These considerations indicate that any suggestion of imminence can be inferred from the utterance interpretation context alone, whereas the overtones of prior intention and inevitability, whilst not directly encoded by *be going to*, are also not purely inferred. This entails that prior intention and inevitability, being derived from a combination of information encoded in *be going to* and inference, are explicatures of utterances in which they arise, where "an explicature is a combination of linguistically encoded and contextually inferred conceptual features" (Sperber & Wilson 1986:182). Imminence, however, is simply inferred from the utterance interpretation context and is therefore an implicature of the utterances in which it is recovered.

*Postscript: the pragmatics of will*

The ideas of intention subsequent to a request and contingency often associated with *will* (in contrast to prior intention and inevitability associated with *be going to*) are always cancellable, suggesting that these are implicatures rather than explicatures:

Can somebody visit John tomorrow?

(6.66) I'll visit him, in fact I was already intending to.

(6.67) a. It's going to explode whether anyone goes near it or not.

b. It will explode whether anyone goes near it or not.

If these overtones are not explicit, that is, not derived from a logical form containing *will*, why do they typically arise from utterances of sentences containing *will* rather than any other form? The answer lies in the principle of relevance. *Will* with future reference contrasts with *be going to* (among other future expressions), and an addressee, aware that a speaker has a choice of various forms for the expression of futurity in English, assumes that the choice of *will* is motivated by some



communicative purpose. Specifically, a speaker who does not want an addressee to recover any of the overtones which can be inferred from an utterance of a sentence containing *be going to* (or some other form) may use *will*. Thus, by inference alone, an addressee may recover intention subsequent to, rather than prior to, a request, or contingency in contrast to inevitability, depending on the utterance interpretation context. In this way, adequate contextual effects are achieved, and the additional inferential processing entailed is offset by the frequency of mention of *will* over all other future expressions.<sup>73</sup> What maintains the integrity of *will*, or indeed of any linguistic sign, is its common core of (often semantically underdetermined) meaning plus "the opposition between that sign and other signs with related but contrasting meanings" (Contini-Morava 1989:39) including, in the case of *will*, other modal auxiliaries and *be going to*.

#### 6.4.4 The conceptual semantic content of *be going to*

When *be going to* is used, the overtones discussed above are not always recovered (as opposed to activated) by an addressee, whereas future time reference is. At the end of §5.3.2 I demonstrated, with respect to *will*, that the procedural information encoded by a gram is always recovered by the addressee (since being procedural it can only reduce processing effort). Any conceptual information also encoded (through semantic retention) is only recovered if this results in adequate contextual effects, thereby justifying the additional processing effort involved in recovering and enriching the additional conceptual information. This suggests that whilst future time reference is procedurally encoded by *be going to* (as I argued in §6.4.1), the overtones of prior intention and inevitability might result from semantic retention. If the overtones do in fact arise as a result of semantic retention, the conceptual semantic content of *be going to* must be specified; this is the task of the current section.

An account of the conceptual information encoded by *be going to* must fulfil the following three criteria. First, the conceptual semantics of *be going to* must be compatible with the semantics of the lexical source construction, since the conceptual information encoded by *be going to* results from semantic retention, whereby "certain more specific semantic nuances of the source construction can be retained in certain contexts long after grammaticization has begun." (Bybee *et al* 1994:16) Second, it must be characterised as a state, rather than as an event, as the co-occurrence of *be going to* with *already* (which imposes constraints on the temporal reference of ongoing or resultant states rather than events) demonstrated (see §6.3.2). Finally, the

---

<sup>73</sup> The Survey of English Usage (University College London 1960) cited in Haegeman (1983:15) recorded 2,668 occurrences of *will* compared with 539 of *be going to*.

conceptual semantic content of *be going to* must be capable of being inferentially enriched to give rise to the overtones of prior intention and inevitability.

Let us take the first criterion, compatibility with the semantics of the lexical source construction. In §5.2.3 I suggested that generalisation was involved in the early semantic development of *be going to*, prior to actual grammaticization (through inference, that is, the conventionalisation of implicature). Generalisation involves the loss of one feature of the meaning of an expression, in this case the notion of physical movement:

<i>be going to</i>		
(i)	physical progression	the subject is in the process of progressing towards a physical goal.
(ii)	general progression	the subject is in the process of progressing towards a goal.

The goal towards which the subject is progressing is a physical location in (i) and the realisation of a further situation or event in (ii). If the subject is in the process of progressing towards the realisation of a further situation or event, the subsequent realisation of that situation or event is implied. The grammaticization of *be going to* as an operator encoding future time reference can then be easily achieved through the conventionalisation of this implicature.

The second criterion, that the conceptual semantics of *be going to* must be characterised as a state, has been fulfilled. The state of being in the process of progressing towards the realisation of a subsequent situation or event is a state having the potential for the subsequent realisation of that event; this simplification gives rise to the following paraphrase of the conceptual semantics of *be going to*:

A SR, SR2, containing both a form of *be going to* and an embedded SR, SR1, (i.e. OPERATOR [2 BE GOING TO [1 ...]]) is a representation of a state having the potential for the subsequent realisation of SR1.

With this definition of the semantics of *be going to*, it is possible to account for the overtones of prior intention and inevitability, thereby fulfilling the third criterion. Prior intention is conveyed by the use of *be going to* in (6.42a):

- Can somebody visit John tomorrow?
- (6.42) a. I'm going to visit him.  
b. I'll visit him.

These can be represented as the following logical forms:

- (6.68) a. NON-PAST [2 BE GOING TO [1 I VISIT HIM]]  
 b. WILL [1 I VISIT HIM]

In (6.68a) SR2 represents a state with the potential for the subsequent realisation of SR1. In (6.68b) SR1 only became a potential event once (6.42b) had been uttered; but in (6.68a) the potential for subsequent realisation of the event [I VISIT HIM] is stated as currently existing, in this case inferred to be a prior intention. The contrast between *be going to* and *will* on this interpretation is illustrated in (6.69) below:

- (6.69) I wasn't going to (visit him) but I will (visit him) if you like.  
 PAST [3 NOT [2 BE GOING TO [1 I VISIT HIM]]] but WILL [1 I VISIT HIM] ...

(6.69) can be paraphrased as follows: in the past there was no potential for the situation [I VISIT HIM] to be subsequently realised, but now (that I have been asked) [I VISIT HIM] is a potential situation.

Similarly, the overtone of inevitability conveyed by (6.50a) arises out of the fact that *be going to* indicates that the potential for the subsequent realisation of the event [IT EXPLODE] is current and actual:

- (6.50) a. Don't go near it! It's going to explode!  
 NON-PAST [2 BE GOING TO [1 IT EXPLODE]]

SR2 is a state having the potential for the subsequent realisation of SR1; from this, and the utterance interpretation context, the overtone of inevitability can be inferred.

## **6.5 Summary**

In this chapter, *will*, *shall* and *be going to* were all characterised as exponents of procedural encoding at phase two, with scope over SRs. Although these three forms are all used to express future time reference, the procedural information encoded by each is distinct. This reflects their differing lexical sources, which also determine any conceptual semantic content also encoded. The semantic and scopal differences between *will*, *shall* and *be going to* are reflected in the differences in distribution and interpretation amongst them which I have discussed. Below, I shall review my analyses of *will*, *shall* and *be going to* in turn.

*Will* encodes information relating SRs in its scope to the irrealis cognitive domain; that is, it indicates that any SR in its scope is to be entertained as unverified, or potential. The temporal reference of a SR within the scope of *will* must either be specified adverbially or inferred from the utterance interpretation context, since temporal reference constitutes no part of the semantics of *will*:

- (6.8)           How old will Alison be next month?  
(6.9)           If Alison was born in April 1970, how old is she?  
(6.7)           She'll be twenty six.

With certain utterances of *will*, especially in interrogative and conditional clauses, volitional overtones may arise:

- (6.70)          Will you do the shopping?  
(6.71)          If you'll do the shopping for me, I'll give you some money.

In chapter 5 I argued that these are the result of semantic retention, reflecting the volitional semantics of the lexical source expression from which *will* the grammatical marker has developed.

*Shall* has developed from a lexical source expressing obligation. This has given rise to procedural information specifying that a SR within the scope of *shall* is to be interpreted as an agent-event (that is, an event brought about by motivated human activity) rather than a world-event (that is, an event brought about independently of motivated human activity). As a result, dynamic and deontic modality can be inferred from the use of *shall*, since these involve factors intrinsically linked with motivated human activity: permission and obligation in the case of deontic modality, volition and ability in the case of dynamic modality. Epistemic modality can not be inferred from the use of *shall*, since this is "an unverified correspondence between a SITUATION REPRESENTATION and either an obtaining or [...] a previous WORLD SITUATION or a future WORLD-EVENT." (Klinge 1993:346) In contrast, *will* can express all three modalities, since it is unspecified with respect to the agent-event versus world-event distinction.

The agent-event/world event distinction is only operative in English with futurate situations and events (thus it only affects the interpretation of clauses with imperfective or unmarked aspect where these have future time reference - see §3.3.3). As a result, the agent-event procedural semantics of *shall* subsumes future time reference. A process of semantic generalisation has begun in *shall*, with the result that

*shall* can express simple futurity with first person subjects. However, this generalisation has not extended to non-first person subjects; this arrested development coincides with a decrease in the frequency of *shall* and a corresponding increase in the frequency of *will* and, in particular, *be going to*.

With respect to the *be going to* construction, I demonstrated that, although it belongs to a different syntactic class from the modal auxiliaries, *be going to* can also be characterised as an operator, although with narrower scope than the modal auxiliaries. In a construction of the form OPERATOR [<sub>2</sub> BE GOING TO [<sub>1</sub> ... ]], *be going to* encodes procedural information to locate SR1 at a time subsequent to that of SR2 (SR1 > SR2). This procedural characterisation reflects the following facts: First, that *be going to* exhibits many of the features of grammatical markers: semantic bleaching, phonetic reduction (in the *gonna* allomorph) and syntactic rigidity (see §6.3.1). Secondly, that the lexical source construction of *be going to* - consisting of a movement verb with imperfective aspect and an allative component - is a typical source construction for future grams (Bybee et al 1994:11). Third, that positing a temporal relation of the form SR1 > SR2 accounted for the unacceptability of (6.28) in contrast to the acceptable (6.31):

(6.28) \* They are going to have arrived two hours ago.

(6.31) They will have arrived two hours ago.

*Be going to* also exhibits semantic retention; the conceptual semantics of *be going to* was paraphrased in §6.4.4 as:

A SR, SR2, containing both a form of *be going to* and an embedded SR, SR1, (i.e. OPERATOR [<sub>2</sub> BE GOING TO [<sub>1</sub> ...]]) is a representation of a state having the potential for the subsequent realisation of SR1.

This conceptual semantic content a) is compatible with the lexical source construction of *be going to*, b) is characterised as a state - reflecting the co-occurrence of *be going to* with *already* (§6.3.2), and c) gives rise to the overtones of prior intention and inevitability. These overtones are explicatures (since they arise with utterances containing *be going to* but not with utterances containing *will*) and are not cancellable:

Can somebody visit John tomorrow?

(6.57) ? I'm going to visit him but I wasn't intending to.

(6.58) Don't go near that parcel! It's going to explode!  
? But if you keep well clear it won't explode.

The overtones of intention subsequent to a request and contingency, associated with *will*, and the overtone of imminence, often associated with *be going to*, on the other hand, are implicatures, being always cancellable:

Can somebody visit John tomorrow?

(6.66) I'll visit him, in fact I was already intending to.

(6.67) b. It will explode whether anyone goes near it or not.

(6.56) What are you going to play us?

(6.65) Some Chopin, but first I'd like to say a few words...

To summarise, I demonstrated that *will*, *shall* and *be going to* encode procedural information allowing these expressions to contribute to (without fully determining) propositions expressed by utterances in which they occur, by constraining the inferential interpretation of SRs within their scope. However, purely procedural characterisations of *will* and *be going to* fail to provide comprehensive accounts of their distribution and interpretation, and I demonstrated that these expressions also encode conceptual information (no longer encoded by the older *shall*).

## 7 GENERAL CONCLUSIONS

The aims of the thesis, as stated in the introduction, were to explore the distinction between conceptual and procedural encoding, and in particular to determine and characterise exponents of procedural encoding constraining the manipulation of sub-propositional conceptual representations. The first two chapters described the theoretical basis of the distinction in terms of a three-phase, relevance theoretic model of utterance interpretation, and reviewed some of the literature concerning it. Then, in chapters 3 and 4, I demonstrated how the distinction could be extended to provide a procedural account of pronouns, and tense, aspect and modality markers. Finally, in chapters 5 and 6, I discussed the relationship between the conceptual/procedural distinction in semantics and grammaticization in tense, aspect and modality markers, and demonstrated how a comprehensive account of future expressions in English must make reference to both conceptual and procedural encoding. I shall discuss my conclusions in greater detail below.

### 7.1 The theoretical basis

The cognitive basis for a distinction between conceptual and procedural information lies in the principle of relevance. The need to achieve adequate contextual effects requires that utterances contain or make reference to information capable of being developed into fully propositional conceptual representations, which can be combined with assumptions in the utterance interpretation context to yield contextual effects; this is conceptual information. Balancing this need to achieve adequate effects is the requirement to do so as efficiently as possible; procedural information reduces processing effort by constraining the utterance interpretation process, thereby increasing cognitive efficiency and contributing to relevance. Since humans are rational information processors who both produce and receive linguistically encoded information, it is to be expected that linguistic systems will encode not only conceptual information, but procedural information also.

According to relevance theory, inferential communication takes precedence over decoding which, being fast, automatic and informationally encapsulated, incurs less processing effort than inference, which involves central cognitive processes. Accordingly, procedural encoding constrains inferential processing. In chapter 1 I introduced a three-phase model of utterance interpretation, consisting of a modular decoding phase and two inferential phases. During phase three of utterance interpretation, fully propositional conceptual representations are manipulated with respect to an utterance interpretation context (i.e. new information is contextualised in

old information) resulting in contextual effects. Blakemore (1987, 1988a) proposed that this process of contextualisation is constrained by discourse connectives such as *so* and *after all* which reduce the hypothesis space to be searched and thereby reduce processing effort. In chapter 2 I developed this theme, and discussed discourse connectives and syntactic constructions which constrain the computation of implicatures, and the Kiswahili interrogative marker, JE, which constrains the computation of higher-level explicatures. Although these are clear cases of procedural encoding, I mentioned the claim made in Groefsema (1992) that a single expression might in theory encode both conceptual and procedural information. Also discussed, was the possibility of procedural encoding functioning at phase two of utterance interpretation, constraining the inferential processes involved in the construction of fully propositional conceptual representations; this extension of the distinction was explored in following chapters.

## 7.2 Procedural encoding at phase two

In chapter 3 I developed Wilson & Sperber's (1993) proposal that pronouns encode procedural information constraining the identification of propositions. I maintained the view of inference as the manipulation of conceptual representations by proposing that pronouns constrain the manipulation of sub-propositional conceptual representations of entities, projected by the logical entries of predicates. I further proposed that pronouns constrain reference assignment on the basis of Ariel's Accessibility Hierarchy and on the basis of other information such as number, gender, case, or noun class. I demonstrated that, while English personal pronouns and Kiswahili Object Markers differed significantly in terms of their syntactic, semantic and discourse functions, they could all be characterised as exponents of procedural encoding at phase two.

I then discussed the role of procedural encoding in the determination of temporal reference, aspectual status, and modality, all of which can be inferentially determined and contribute to the construction of propositional conceptual representations. I demonstrated that markers of tense and aspect cannot be characterised in directly truth-conditional terms since, while they contribute to propositions, they do not fully determine them - as inference is still integral to the determination of temporal and aspectual information (in the absence of explicit temporal and aspectual adverbials), even when markers of tense and aspect are present. I proposed a characterisation of tense and aspect markers as 'operators', that is, exponents of procedural encoding which constrain the inferential manipulation of sub-propositional conceptual representations of situations and events ('situation representations' or SRs for short).



Chapter 4 consisted of a more detailed analysis of the English modal auxiliaries *can*, *may*, *must* and *should* in procedural terms (based largely on Groefsema 1995a). Although this purely procedural analysis accounted for their distributions and interpretations more successfully than the alternatives, it failed to account for certain features of the semantics of *should* and *must* in particular, and did not deal with variations in usage. Variation is a feature of linguistic change, so in the following chapter I turned my attention to grammaticization.

### 7.3 A diachronic perspective

The final two chapters introduced a diachronic and variationist perspective into relevance theory, through relating conceptual and procedural encoding to grammaticization and the analysis of English future expressions, taking account of variation and historical factors.

In chapter 5 I explored the relation between conceptual/procedural distinction at phase two and grammaticization, with respect to the development of tense, aspect and modality markers. I discussed three facets of grammaticization from a relevance theoretic perspective: the question of whether grammaticization is semantically gradual or instantaneous, the causes of grammaticization, and the phenomenon of semantic retention (where semantic overtones of the lexical source construction of a grammatical marker are recoverable in certain contexts). I first showed how the model of the conceptual/procedural distinction developed in the thesis, which views it as a dichotomy excluding any intermediate information type(s), was compatible with Givón's (1991) findings that the onset of grammaticization is semantically instantaneous and does not involve any intermediate stage between lexical and (functionally) grammatical expression. I then discussed the mechanisms of semantic change involved in the onset of grammaticization, and concluded that the conventionalisation of implicature (an inferential process) was arguably the only sure semantic mechanism involved in the development of lexical expressions into grammatical markers. This was to be expected, given that grammatical markers had been characterised as exponents of procedural information, the purpose of which is to constrain the inferential processes of utterance interpretation. Finally, I discussed semantic retention, and showed how a monosemous account of a grammatical marker exhibiting semantic retention was possible if the gram were characterised as an exponent of both conceptual and procedural encoding (as proposed earlier). The conditions under which *will* exhibits semantic retention were accounted for by reference to the principle of relevance.

I concluded by providing an account of the semantics of *will*, *shall* and *be going to*, based on the conceptual/procedural distinction and taking account of the historical development of these expressions and their current variation in usage. I demonstrated that while they are all operators, *will* and *be going to* also exhibit semantic retention and can therefore only be accurately characterised if conceptual encoding in these forms is taken into account.

#### 7.4   Prospects

In the postface to the second edition of *Relevance* Sperber & Wilson write:

“It may turn out that the conceptual/procedural distinction will shed more light on linguistic semantics than the traditional distinction between truth-conditional and non-truth-conditional meaning.”

(S&W 1995:259)

I hope that in this thesis I have shed some light on the conceptual/procedural distinction, and that a sense of its potential importance has been conveyed. At phase three of utterance interpretation, procedural encoding can be found not only in discourse connectives, but in syntax (where much work remains to be done relating procedural encoding and the various manifestations of information packaging) and in the interpretive use of language for the expression of higher-level explicatures. By extending the distinction to incorporate inferential processing at phase two, it may shed light on a whole range of phenomena. The notion of procedural encoding has implications not only for pronouns but also for the study of reference assignment and anaphora resolution in general. In the study of tense, aspect and modality also, procedural analyses promise exciting alternatives to directly truth-conditional accounts.

Much work remains to be done, but I hope that in the three-phase model of utterance interpretation and in the characterisation of procedural information in terms of constraints on the manipulation of sub-propositional and propositional conceptual representations, I have provided useful tools for this task.

## Key to Kiswahili sources

- HA** Anonymous 1966 *Hekaya za Abunuwas na hadithi nyingine* London, Macmillan and Co.
- Kiu** Mohamed, M. S. 1975 *Kiu* Dar es Salaam, East African Publishing House
- KMF** Shafi, S. A. 1978 *Kasri ya Mwinyi Fuad* Dar es Salaam, Tanzania Publishing House
- MWK** Abdullah, M. S. 1960 *Mzimu wa Watu wa Kale* Nairobi, Kenya Literature Bureau
- NS** Msokile, M. 1981 *Nitakuja kwa Siri* Dar es Salaam, Dar es Salaam University Press
- RM** Kezilahabi, E. 1971 (2nd edition 1988) *Rosa Mistika* Dar es Salaam, Dar es Salaam University Press
- Uk** Said, S. 1979 *Ukombozi* Dar es Salaam and Arusha, Eastern Africa Publications
- WU** Hussein, E. 1971 *Wakati Ukuta* Nairobi, East Africa Publishing House

## Bibliography

- Allan, K. 1983 'Anaphora, cataphora and topic focusing functions of the object prefix in Swahili' In Dihoff, I. R. (ed) *Current Approaches to African Linguistics vol. 1* Dordrecht, Foris Pp.322-335
- Altmann, G. T. M., Garnham, A. and Dennis, Y. 1992 'Avoiding the garden path: Eye movements in context' *Journal of Memory and Language* 31:685-712
- Altmann, G. T. M. and Steedman, M. 1988 'Interaction with context during human sentence processing' *Cognition* 30:191-238
- Anderson, A. R. and Belnap, N. D. Jr. 1975 *Entailment Vol. 1* Princeton, Princeton University Press
- Ariel, M. 1988 'Referring and accessibility' *Journal of Linguistics* 24:65-87
- Ariel, M. 1990 *Accessing NP antecedents* London, Routledge
- Ariel, M. 1991 'The function of accessibility in a theory of grammar' *Journal of Pragmatics* 16:443-463
- Ariel, M. 1994 'Interpreting anaphoric expressions: a cognitive versus a pragmatic approach' *Journal of Linguistics* 30:3-42
- Ashton, E. O 1947 *Swahili Grammar* 2nd edition London, Longman
- Bánrétí, Z. 1994 'Modularity in Sentence Parsing: Grammaticality Judgements by Broca's Aphasics' *Working Papers in the Theory of Grammar* Vol.1, No.4 Budapest, Hungarian Academy of Sciences, Research Institute for Linguistics
- Barense, D. B. 1980 *Tense Structure and Reference: a first-order non-modal analysis* Bloomington, Indiana University Linguistics Club
- Barwise, J. and Perry, J. 1983 *Situations and Attitudes* Cambridge MA, MIT Press
- Bickerton, D. 1975 *Dynamics of a Creole System* Cambridge, C.U.P.
- Bickerton, D. 1981 *Roots of Language* Ann Arbor, Karoma
- Bickerton, D. 1984 'The language bioprogram hypothesis' *The Behavioral and Brain Sciences* 7:173-222

- Bickhard, M. H. and Campbell, R. L. 1992 'Some foundational questions concerning language studies: With a focus on categorial grammars and model-theoretic possible world semantics' Journal of Pragmatics 17:401-433
- Binnick, R. I. 1972 '*Will and be going to II*' In Papers from The 8th Regional Meeting Chicago Linguistic Society
- Blakemore, D. 1987 *Semantic Constraints on Relevance* Oxford, Basil Blackwell
- Blakemore, D. 1988a "'So" as a constraint on relevance' In Kempson (ed) Pp.183-195
- Blakemore, D. 1988b 'The organization of discourse' In Newmeyer (ed) Vol. IV Pp.229-250
- Blakemore, D. 1990 'Constraints on interpretations' In Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society: Parasession on the Legacy of Grice Berkeley, CA, Berkeley Linguistics Society Pp. 363-371
- Blass, R. 1990 *Relevance relations in discourse: A study with special reference to Sissala* Cambridge, Cambridge University Press
- Bock, J. K. and Loebell, H. 1990 'Framing sentences' Cognition 35:1-39
- Bock, J. K., Loebell, H. and Morey, R. 1992 'From conceptual roles to semantic relations: bridging the syntactic cleft' Psychological Review 99:150-171
- Bradshaw, J. 1982 *Word Order Change in Papua-New Guinea Austronesian Languages* University of Hawaii PhD thesis
- Bresnan, J. and Mchombo, S. 1987 *Topic, Pronoun and Agreement in Chichewa* Stanford CA, CSLI
- Bybee, J. L. 1985 *Morphology: A Study of the Relation between Meaning and Form* Amsterdam, John Benjamins
- Bybee, J. L. 1988 'Semantic substance vs. contrast in the development of grammatical meaning' In Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society: Parasession on Grammaticalization Pp.247-264
- Bybee, J. L., Pagliuca, W. and Perkins, R. D. 1991 'Back to the future' In Traugott, E. C. and Heine, B. (eds) *Vol.II: Focus on Types of Grammatical Markers* Pp.17-58
- Bybee, J. L., Perkins, R. D. and Pagliuca, W. 1994 *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World* Chicago/London, University of Chicago Press
- Carey, K. 1990 'The role of conversational implicature in the early grammaticalization of the English perfect' In Proceedings of the Sixteenth Annual Meeting of the Berkeley Linguistics Society: Parasession on the Legacy of Grice Berkeley, CA, Berkeley Linguistics Society Pp. 371-380
- Carston, R. 1988a 'Language and cognition' In Newmeyer (ed) Vol. III Pp.38-68
- Carston, R. 1988b 'Implicature, explicature and truth-theoretic semantics' In Kempson (ed) Pp.155-181
- Carston, R. 1993 'Conjunction, explanation and relevance' Lingua 90:27-48
- Chomsky, N. 1980 *Rules and Representations* New York, Columbia University Press
- Chomsky, N. 1991 'Some notes on economy of derivation and representation' In Freidin, R. (ed) *Principles and Parameters in Comparative Grammar* Cambridge MA, MIT Press Pp.417-545
- Chomsky, N. 1993 'A minimalist program for linguistic theory' In Hale, K. and Keyser, S.J. (eds) *The View from Building 20* Cambridge MA, MIT Press Pp1-52
- Chomsky, N. 1995 'Bare phrase structure' In Webelhuth, G. (ed) *Government and Binding Theory and the Minimalist Programme* Oxford, Blackwell Pp.383-440

- Chung, S. and Timberlake, A. 1985 'Tense, Aspect and Mood' In Shopen, T. (ed) *Language Typology and Syntactic Description - Vol.III: Grammatical Categories and the Lexicon* Cambridge, C.U.P. Pp.202-258
- Clark, H. H. 1982 reply to Sperber and Wilson 1982 in Smith (ed) Pp.124-127
- Coates, J. 1983 *The Semantics of the Modal Auxiliaries* London etc., Croom Helm
- Cobb, M. 1995 'Vowel Harmony in Zulu: The Interaction of Licensing Constraints and Constituent Structure' Paper read at the 4th Manchester Postgraduate Linguistics Conference, 11/3/1995
- Comrie, B. 1981 *Language Universals and Linguistic Typology* Chicago, University of Chicago Press
- Comrie, B. 1985 *Tense* Cambridge etc., C.U.P.
- Contini-Morava. E. 1989 *Discourse Pragmatics and Semantic Categorization: The Case of Negation and Tense-Aspect with Special Reference to Swahili* Berlin/New York, Mouton de Gruyter
- Crain, S. and Steedman, M. 1985 'On not being led up the garden path: the use of context by the psychological syntax processor' In Dowty *et al* (eds) Pp.320-358
- Crowley, T. 1987 'Serial verbs in Paamese' Studies in Language 11:35-84
- Culler, J. 1976 *Saussure* Glasgow, Fontana/Collins
- Dahl, Ö. 1985 *Tense and Aspect Systems* Oxford, Basil Blackwell
- Dahl, Ö. 1987 'Comrie's *Tense*' Folia Linguistica 21:489-502
- Davidson, D. 1979 'Moods and Performances' In Davidson, D. 1984 *Inquiries into Truth and Interpretation* Oxford, Clarendon Press Pp.109-121
- Declerck, R. 1984 'Pure future' *will* in *if*-clauses' Lingua 63:279-312
- Declerck, R. 1989 *Tense in Discourse* Bloomington, Indiana University Linguistics Club
- Declerck, R. 1991 *Tense in English: Its Structure and Use in Discourse* London, Routledge
- Declerck, R. and Depraetere, I. 1995 'The double system of tense forms referring to future time' Journal of Semantics 12:269-310
- Depraetere, I. 1994 'Some observations on the expression of temporal relations in future-time relative clauses' Linguistics 32:459-473
- Downing, A. and Locke, P. 1992 *A University Course in English Grammar* New York etc., Prentice Hall
- Dowty, D. R. 1977 'Toward a semantic analysis of verb aspect and the English 'imperfective' progressive' Linguistics and Philosophy 1:45-77
- Dowty, D., Karttunen, L. and Zwicky, A. (eds) 1985 *Natural Language Processing* Cambridge, C.U.P.
- Driever, D. 1976 *Aspects of a Case Grammar of Mombasa Swahili* Hamburg, Helmut Buske Verlag
- Duranti, A. 1979 'Object clitic pronouns in Bantu and the topicality hierarchy' Studies in African Linguistics 10:31-45
- Durst-Andersen, P. 1986 'Russian qualitative aspect: referential and propositional semantics' In Sørensen, F. (ed) *Aspects of Aspect* Copenhagen, Cebal Pp.27-57
- Durst-Andersen, P. 1992 *Mental Grammar: Russian Aspect and Related Issues* Columbus OH, Slavica

- Ehrman, M. E. 1966 *The Meanings of the Modals in Present-day American English* The Hague, Mouton
- Etchemendy, J. 1988 'Models, semantics and logical truth' Linguistics and Philosophy 11:91-106
- Fodor, J. A. 1983 *The Modularity of mind* Cambridge MA, M.I.T. Press
- Fodor, J. A. 1985 'Précis of *The Modularity of mind*' The Behavioral and Brain Sciences 8:1-5
- Foster-Cohen, S. H. 1994 'Exploring the boundary between syntax and pragmatics: relevance and the binding of pronouns' Journal of Child Language 21:237-255
- Frazier, L. 1979 *On Comprehending Sentences: Syntactic Parsing Strategies* Bloomington, Indiana University Linguistics Club
- Frege, G. 1892 'Über Sinn und Bedeutung' Zeitschrift für Philosophie und philosophische Kritik 100:25-50
- Fries, C. C. 1927 'The expression of the future' Language 3:87-95
- Garrod, S., Freudenthal, D. and Boyle, E. 1994 'The role of different types of anaphor in the on-line resolution of sentences in a discourse' Journal of Memory and Language 33:39-68
- Gazdar, G. 1979 *Pragmatics: implicature, presupposition and logical form* New York, Academic Press
- Gazdar, G. and Good, D. 1982 reply to Sperber and Wilson 1982 in Smith (ed) Pp.88-100
- Gazdar, G., Klein, E., Pullum, G. K. and Sag, I. A. 1985 *Generalized Phrase Structure Grammar* Cambridge MA, Harvard University Press
- Gibbs, R. W. 1986a 'On the psycholinguistics of sarcasm' Journal of Experimental Psychology 115:3-15
- Gibbs, R. W. 1986b 'Skating on thin ice: Literal meaning and understanding idioms in conversation' Discourse Processes 9:17-30
- Gibbs, R. W. 1989 'Understanding and literal meaning' Cognitive Science 13:243-51
- Givón, T. 1975a 'Serial verbs and syntactic change: Niger-Congo' In Li, C. N. (ed) *Word Order and Word Order Change* Austin, University of Texas Press Pp.47-112
- Givón, T. 1975b 'Focus and the scope of assertion' Studies in African Linguistics 6:185-205
- Givón, T. 1976 'Topic, pronoun and grammatical agreement' In Li, C. N. (ed) *Subject and Topic* New York, Academic Press Pp.151-188
- Givón, T. (ed) 1979 *Syntax and Semantics 12: Discourse and Syntax* New York etc. Academic Press
- Givón, T. 1984 *Syntax: A Functional-Typological Introduction* Amsterdam/Philadelphia, John Benjamins
- Givón, T. 1991 'Serial verbs and the mental reality of 'event': grammatical vs. cognitive packaging' In Traugott, E. C. and Heine, B. (eds) *Vol.1: Focus on Theoretical and Methodological Issues* Pp.81-127
- Gorayska, B. and Lindsay, R. 1993 'The roots of relevance' Journal of Pragmatics 19:301-323
- Gordon, P. C. and Chan, D. 1995 'Pronouns, passives, and discourse coherence' Journal of Memory and Language 34:216-231
- Grice, H. P. 1957 'Meaning' Philosophical Review 66:377-388

- Grice, H. P. 1989 *Studies in the Way of Words* Cambridge MA, Harvard University Press
- Groefsema, M. 1992 *Processing for Relevance: A pragmatically based account of how we process natural language* unpublished PhD thesis, University College London
- Groefsema, M. 1995a 'Can, may, must and should: A Relevance theoretic account' Journal of Linguistics 31:53-79
- Groefsema, M. 1995b 'Understood arguments: A semantic/pragmatic approach' Lingua 96:139-161
- Haack, S. 1978 *Philosophy of Logics* Cambridge etc., C.U.P.
- Haegeman, L. 1983 *The Semantics of will in present-day British English: A Unified Account* Brussels, Paleis der Academiën
- Haegeman, L. 1989 'Be going to and will: a pragmatic account' Journal of Linguistics 25:291-317
- Haegeman, L. and Wekker, H. Chr. 1984 'The syntax and interpretation of futurate conditionals in English' Journal of Linguistics 20:45-55
- Halliday, M. A. K. 1967 'Notes on transitivity and theme in English, part II' Journal of Linguistics 3:199-244
- Halliday, M. A. K. and Hasan, R. 1976 *Cohesion in English* London, Longman
- Handley, S. and Buck, E. 1995 'Representation and relevance in human reasoning' Paper presented at the University of Hertfordshire Relevance Theory Workshop 27-29 October 1995
- Harman, G. and Davidson, D. (eds) 1972 *Semantics of Natural Language* Reidel
- Hatav, G. 1993 'The aspect system in English: an attempt at a unified analysis' Linguistics 31:209-237
- Hauner, M. 1984 'Complex Tenses and Style in the Novels of E. Kezilahabi' In Maw, J. and Parkin, D. (eds) *Swahili Language and Society* Wien, Universität Wien Pp.107-125
- Hausser, R. 1989 *Computation of Language* Berlin etc., Springer-Verlag
- Hawkins, J. 1978 *Definiteness and Indefiniteness: A Study in Reference and Grammaticality Prediction* London, Croom Helm
- Heim, I. 1983 'File Change Semantics and the familiarity theory of definiteness' In Bäuerle, R. et al (eds) *Meaning, Use, and Interpretation of Language* Berlin/New York, Mouton de Gruyter Pp.164-189
- Heine, B., Claudi, U. and Hünemeyer, F. 1991 'From cognition to grammar - evidence from African Languages' In Traugott, E. C. and Heine, B. (eds) *Vol. I: Focus on Theoretical and Methodological Issues* Pp.149-187
- Hermerén, L. 1978 *On modality in English: a study in the semantics of the modals* Lund, Gleerup
- Higginbotham, J. 1988 'Contexts, models, and meanings: a note on the data of semantics' In Kempson (ed) Pp.29-48
- Hinnebusch, T. J. 1979 'Swahili' In Shopen, T. (ed) *Languages and Their Status* Cambridge MA, Winthrop, Pp.209-293
- Hintikka, J. 1969 *Models for Modalities* Reidel
- Hirst, G. 1987 *Semantic interpretation and the resolution of ambiguity* Cambridge etc., C.U.P.

- Hirtle, W. H. and Curat, V. N. 1986 'The simple and the progressive: 'future' use' Transactions of the Philological Society 1986:42-84
- Hopper, P. J. 1991 'On some principles of grammaticization' In Traugott, E. C. and Heine, B. (eds) *Vol.1: Focus on Theoretical and Methodological Issues* Pp.17-35
- Hopper, P. J. and Traugott, E. C. 1993 *Grammaticalization* Cambridge, C.U.P.
- Hyman, L. and Duranti, A. 1982 'On the object relation in Bantu' In Hopper, P. and Thompson, S. (eds) *Syntax and Semantics 15: Studies in Transitivity* New York, Academic Press Pp.217-239
- Jackendoff, R. S. 1983 *Semantics and Cognition* Cambridge MA, MIT Press
- Jackendoff, R. S. 1990 *Semantic Structures* Cambridge MA, MIT Press
- Jardine, N. 1975 'Model theoretic semantics and natural language' In Keenan, E. L. (ed) *Formal Semantics of Natural Language* Cambridge, C.U.P. Pp.219-240
- Johnson, F. 1939 *A Standard Swahili-English Dictionary* Nairobi and Dar es Salaam, Oxford University Press
- Johnson-Laird, P. N. 1983 *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness* Cambridge, C.U.P.
- Jucker, A. H. 1993 'The discourse marker *well*': A relevance-theoretical account' Journal of Pragmatics 19:435-452
- Kamp, H. 1981 'A theory of truth and semantic representation' In Groenendijk, J., Janssen, T. and Stokhof, M. (eds) *Formal Methods in the Study of Language, Part 1* Amsterdam, Mathematisch Centrum Pp.277-322
- Kamp, H. and Reyle, U. 1993 *From Discourse to Logic* Dordrecht etc., Kluwer Academic
- Kapinga, M. C. 1983 *Sarufi Maumbo ya Kiswahili Sanifu* Dar es Salaam, Dar es Salaam University, Taasisi ya Uchunguzi wa Kiswahili
- Kaplan, D. 1989 'Demonstratives' In Almog, J., Perry, J., and Wettstein, H. (eds) *Themes from Kaplan* Pp.481-563 Oxford, O.U.P.
- Keenan, E. O. 1976 'On the universality of conversational implications' Language in Society 5:67-80
- Kempson, R. M. (ed) 1988 *Mental Representations: The interface between language and reality* Cambridge, C.U.P.
- Kempson, R. M. 1988a 'Grammar and conversational principles' In Newmeyer (ed) Vol. II Pp.139-163
- Kempson, R. M. 1988b 'The relation between language, mind, and reality' In Kempson (ed) Pp.3-25
- Kempson, R. M. 1988c 'Logical form: the grammar-cognition interface' Journal of Linguistics 24:393-431
- Kempson, R. M. 1994a 'Ellipsis as Labelled Natural Deduction' paper presented at the International Colloquium on the Processes of Linguistic Communication, Bangor 24 June 1994
- Kempson, R. M. 1994b 'Semantics, pragmatics and natural language interpretation' SOAS Working Papers in Linguistics and Phonetics 4:139-168
- Khamisi, A. M. 1972 *Consideration of Grammatical Roles in Swahili* University of York M.Phil. dissertation



- Kihm, A. 1995 'Tayo, the strange bird from New Caledonia: Determiners and tense-aspect in Tayo and their implications for creolization theories' Journal of Pidgin and Creole Languages 10:225-252
- Klinge, A. 1993 'The English modal auxiliaries: from lexical semantics to utterance interpretation' Journal of Linguistics 29:315-357
- Kokolakis, A. 1994 'Pronouns and Relevance Theory' Unpublished MA Dissertation, University College London
- König, E. 1991 *The Meaning of Focus Particles: a comparative perspective* London/New York, Routledge
- Kratzer, A. 1977 'What *must* and *can* must and can mean' Linguistics and Philosophy 1:337-355
- Kratzer, A. 1981 'The notional category of modality' In Eikmeyer, H.-J. and Rieser, H. (eds) *Words, Worlds and Contexts* Berlin, Walter de Gruyter
- Kripke, S. 1972 'Naming and necessity' In Harman, G. and Davidson, D. (eds)
- Lakatos, I. 1978 *The Methodology of Scientific Research Programmes* Cambridge, C.U.P.
- Langacker, R. W. 1983 *Foundations of Cognitive Grammar* Bloomington, Indiana University Linguistics Club
- Leech, G. N. 1971 Meaning and the English Verb (1st edition) London, Longman
- Leech, G. N. 1987 Meaning and the English Verb (2nd edition) London, Longman
- Le Pore, E. 1983 'What model-theoretic semantics cannot do' Synthese 54:167-187
- Leuscher, T. 1995 'Ever and Universal Quantifiers of Time: Observations from Some Germanic Languages' Paper read at the First International Conference in Contrastive Semantics and Pragmatics, University of Brighton, 9 April 1995
- Lewis, C. I. 1918 *A Survey of Symbolic Logic* California University Press
- Lewis, D. 1972 'General semantics' In Harman, G. & Davidson, D. (eds) Pp.169-218
- Lewis, D. K. 1973 *Counterfactuals* Oxford, Blackwell
- Li, C. N. and Thompson, S A. 1979 'Third person pronouns and zero-anaphora in Chinese discourse' In Givón, T. (ed) Pp.311-335
- Lichtenberk, F. 1990 'On the gradualness of grammaticalization' In Traugott, E. C. and Heine, B. (eds) *Vol.1: Focus on Theoretical and Methodological Issues* Pp.37-80
- Lightfoot, D. 1979 *Principles of Diachronic Syntax* Cambridge, C.U.P.
- Lightfoot, D. 1991 *How to Set Parameters: Arguments from Language Change* Cambridge, MA, MIT Press
- Linsky, L. (ed) 1971 *Reference and Modality* Oxford, Oxford University Press
- Luscher, J.-M. 1994 'Les marques de connexion: des guides pour l'interprétation' In Moeschler *et al* 1994 Pp.175-227
- Lyons, J. 1977 *Semantics* Cambridge, C.U.P.
- MacDonald, M. 1993 'The interaction of lexical and syntactic ambiguity' Journal of Memory and Language 32:692-715
- Marslen-Wilson, W. D. and Tyler, L. K. 1987 'Against Modularity' In Garfield, J. L. (ed) *Modularity in Knowledge Representation and Natural-Language Understanding* Cambridge MA/London, MIT Press Pp.37-62

- Marslen-Wilson, W. D., Tyler, L. K. and Koster, C. 1993 'Integrative processes in utterance resolution' Journal of Memory and Language 32:647-666
- Marslen-Wilson, W. D., Tyler, L. K., Warren, P., Grenier, P. and Lee, C. S. 1992 'Prosodic effects in minimal attachment' The Quarterly Journal of Experimental Psychology 45A(1):73-88
- Matsui, T. 1995 *Bridging and Relevance* unpublished PhD thesis, University College London
- Matthews, R. 1989 'Reference Time and Modality' In Schopf, A. (ed) *Essays on Tensing in English, Vol II: Time, Text and Modality* Tübingen, Niemeyer Pp.189-246
- Maw, J. 1976 'Focus and the morphology of the Swahili verb' Buletin of the School of Oriental and African Studies 39:389-402
- Maw, J. and Kelly, J. 1975 *Intonation in Swahili* London, SOAS
- McCawley, J. D. 1971 'Tense and Time Reference in English' In Fillmore, C. J. and Langendoen, D. T. (eds) *Studies in Linguistic Semantics* New York etc., Holt, Rinehart and Winston Pp.97-113
- McIntosh, A. 1966 'Predictive Statements' In Bazell, C. E., Catford, J. C., Halliday, M. A. K. and Robins, R. H. (eds) *In Memory of J. R. Firth* London, Longmans Pp.303-320
- McMahon, A. M. S. 1994 *Understanding Language Change* Cambridge, C.U.P.
- Meillet, A. 1912 'L'évolution des formes grammaticales' In Meillet, A. *Linguistique Historique et Linguistique Générale* Paris, Champion Pp.131-148
- Meinhof, C. 1906 *Gründzüge einer vergleichenden Grammatik der Bantusprachen* Berlin, Reimer
- Mey, J. L. and Talbot, M. 1988 'Computation and the soul' Journal of Pragmatics 12:743-789
- Michaelis, L. A. 1996 'On the use and meaning of *already*' Linguistics & Philosophy 19:477-502
- Mithun, M. 1992 'Is basic word order universal?' In Payne, D. L. (ed.) *Pragmatics of Word Order Flexibility* Amsterdam/Philadelphi, John Benjamins Pp.15-61 (First published in 1987 in the Proceedings of the conference on Coherence and Grounding in Discourse held at the University of Oregon in 1984)
- Moeschler, J. 1994 'Anaphore et déixis temporelles: sémantique et pragmatique de la référence temporelle' In Moeschler *et al* 1994 Pp.39-104
- Moeschler, J., Reboul, A., Luscher, J-M., and Jayez, J. 1994 *Langage et Pertinence* Nancy, Presses Universitaires de Nancy
- Montague, R. 1974 *Formal Philosophy* New Haven CT, Yale University Press
- Moshi, L. P. 1985 *Grammatical Relations in Swahili Discourse* Ann Arbor MI, University Microfilms International
- Muysken, P. 1981 'Creole tense/mood/aspect systems: the unmarked case?' In Muysken, P. (ed) *Generative Studies on Creole Languages* Dordrecht, Foris Pp.181-199
- Myhill, J. 1994 'Intuitive meanings and technical definitions: An analysis of *will* and *gonna*.' University of Michigan Report
- Napoli, D. J. 1993 *Syntax: Theory and Problems* New York/Oxford, O.U.P
- Nemo, F. 1988 'Review of *Relevance: Communication and Cognition*' Journal of Pragmatics 12:791-795
- Newmeyer, F. J. (ed) 1988 *Linguistics: The Cambridge Survey* Cambridge, C.U.P.

- Ni, W., Crain, S. and Shankweiler, D. 1996 'Sidestepping garden paths: Assessing the contributions of syntax, semantics and plausibility in resolving ambiguities' Language and Cognitive Processes 11:283-334
- Nicol, J. L. and Pickering, M. J. 1993 'Processing syntactically ambiguous sentences: Evidence from semantic priming' Journal of Psycholinguistic Research 22:207-237
- Nicol, J. L. and Swinney, D. 1989 'The role of structure in coreference assignment during sentence comprehension' Journal of Psycholinguistic Research 18:5-19
- Nicolle, S. 1995 'In defence of Relevance Theory: A belated reply to Gorayska and Lindsay, and Jucker' Journal of Pragmatics 23:677-681
- Nurse, D. 1989 'Change in Tense and Aspect: Evidence from Noertheast Coastal Bantu Languages' In Haik, I. and Tuller, L. (eds) *Current Approaches to African Linguistics, Vol.6* Dordrecht, Foris Pp.277-297
- Nurse, D. and Hinnebusch, T. J. 1993 *Swahili and Sabaki: A Linguistic History* Berkeley etc., University of California Press
- O'Neill, J. 1988 'Relevance and pragmatic inference' Theoretical Linguistics 15:241-261
- Palmer, F. R. 1979 *Modality and the English Modals* London, Longman
- Palmer, F. R. 1986 *Mood and Modality* Cambridge etc., C.U.P.
- Palmer, F. R. 1990 *Modality and the English Modals* (2nd edition) London/New York, Longman
- Papafragou, A. 1995 'Modality and the semantics/pragmatics interface: the case of *may*, *can*, and *must*' Paper read at University College London, 15 Nov. 1995
- Pawley, A. 1987 'Encoding events in Kalam and English: Different logics for reporting experience' In Tomlin, R. S. (ed) *Coherence and Grounding in Discourse* Amsterdam/philadelphia, John Benjamins Pp.329-360
- Partee, B. 1984 'Nominal and temporal anaphora' Linguistics and Philosophy 7:243-286
- Perkins, M. R. 1982 'The core meanings of the English modals' Journal of Linguistics 18:245-273
- Perkins, M. R. 1983 *Modal Expressions in English* London, Francis Pinter
- Polomé, E. C. 1967 *Swahili Language Handbook* Washington DC, Centre for Applied Linguistics
- Pomerantz, A. 1984 'Agreeing and disagreeing with assessments: Some features of preferred/dispreferred turn shapes' In Maxwell Atkinson, J. and Heritage, J. (eds) *Structures of Social Action: Studies in conversational analysis* Cambridge, C.U.P. Pp.57-101
- Poplack, S. and Tagliamonte, S. 1995 'It's Black and White: the FUTURE in rural Nova Scotia' Unpublished manuscript
- Pylyshyn, Z. (ed) 1987 *The Robot's Dilemma: The frame problem in artificial intelligence* Norwood NJ, Ablex
- Quine, W. V. O. 1961 'Reference and modality' In Quine, W. V. O. 1961 *From a Logical Point of View* Cambridge, MA, Harvard University Press Pp.139-159
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. 1985 *A Comprehensive Grammar of the English Language* London, Longman
- Reboul, A. 1994 'L'anaphore pronominale: Le problème de l'attribution des référents' In Moeschler *et al* 1994 Pp.105-173
- Recanati, F. 1989 'The pragmatics of what is said' Mind and Language 4:295-329

- Reinhart, T. 1983 *Anaphora and Semantic Interpretation* London, Croom Helm
- Reyle, U. and Gabbay, D. M. 1994 'Direct deductive computation on Discourse Representation structures' Linguistics and Philosophy 17:343-390
- Reynolds, K. H. and Eastman, C. M. 1989 'Morphologically based agreement in Swahili' Studies in African Linguistics 20:63-77
- Roberts, C. 1989 'Modal subordination and pronominal anaphora in discourse' Linguistics and Philosophy 12:683-721
- Rugemalira, J. M. 1991 'What is a symmetrical language? Multiple object constructions in Bantu' Proceedings of the 17th Annual Meeting of the Berkeley Linguistics Society: Special Session on African Language Structures Berkeley Linguistics Society Pp.200-209
- Russell, J. 1985 'Swahili quasi-passives: The question of context' In Goyvaerts, D. L. (ed) *African Linguistics* Amsterdam, John Benjamins
- Russell, J. and Rajabu, R. 1995 'Bilingual writers and the syntactic elaboration of Swahili' African Languages and Cultures 8:43-59
- Sanford, A., Moar, K. and Garrod, S. 1988 'Proper names as controllers of discourse focus' Language and Speech 31:43-56
- Sanders, R. E. 1988 'Review of *Relevance: Communication and Cognition*' Language in Society 17:604-609
- Sapir, E. 1912 'The Takelma language of southwestern Oregon' In Boas, F. (ed) *Handbook of American Indian languages, Part 2* Washington DC, Government Printing Office (Bureau of American Ethnology Bulletin 40, 2) Pp.1-296
- Seidl, A. and Dimitriadis, A. 1996 'The discourse function of object marking in Swahili' Unpublished manuscript, University of Pennsylvania
- Smith, N. V. (ed) 1982 *Mutual Knowledge* London etc., Academic Press
- Smith, N. V. 1990 'Observations on the pragmatics of tense' UCL Working Papers in Linguistics 2:82-93
- Smith, N. V. 1993 'Observations sur la pragmatique des temps' Langages 112:26-38
- Smith, N. V. and Wilson, D. 1992 'Introduction to Special Issue on "Relevance Theory"' Lingua 87:1-9
- Sperber, D. 1985 'Apparently irrational beliefs' In Sperber, D. *On Anthropological Knowledge* Cambridge, C.U.P. Pp. 35-63
- Sperber, D. 1994 'The modularity of thought and the epidemiology of representations' In Hirschfeld, L. A. and Gelman, S. A. (eds) *Mapping the Mind: Domain Specificity in Cognition and Culture* Cambridge, C.U.P. Pp.39-67
- Sperber, D., Cara, F. and Girotto, V. 1995 'Relevance theory explains the selection task' Cognition 57:31-95
- Sperber, D. and Wilson, D. 1981 'Irony and the Use-Mention Distinction' In Cole, P. (ed) *Radical Pragmatics* New York etc., Academic Press Pp.295-318
- Sperber, D. and Wilson, D. 1982 'Mutual Knowledge and Relevance in theories of Comprehension' In Smith, N. V. (ed) Pp.61-131
- Sperber, D. and Wilson, D. 1985/6 'Loose talk' Proceedings of the Aristotelian Society 86:153-171
- Sperber, D. and Wilson, D. 1986 *Relevance: Communication and Cognition* Oxford, Basil Blackwell
- Sperber, D. and Wilson, D. 1987 'Précis of *Relevance: Communication and Cognition*' The Behavioral and Brain Sciences 10:697-754

- Sperber, D. and Wilson, D. 1995 *Relevance: Communication and Cognition* (2nd edition) Oxford, Basil Blackwell
- Stalnaker, R. 1974 'Pragmatic presupposition' In Munitz, M. and Unger, P. (eds) *Semantics and Philosophy: Studies in contemporary philosophy* New York, New York University Press Pp.197-213
- Stevenson, R. J. 1993 *Language, Thought and Representation* Chichester etc., Wiley
- Stroik, T. S. 1988 *The Pragmatics of Metaphor* Bloomington, Indiana University Linguistics Club
- Sweetser, E. E. 1982 'Root and epistemic modals: Causality in two worlds' Berkeley Linguistics Society Papers, 484-507
- Sweetser, E. E. 1989 *From Etymology to Pragmatics: the mind-as-body metaphor in semantic structure and semantic change* Cambridge, C.U.P.
- Swinney, D. A. 1979 'Lexical access during sentence comprehension: (Re)considerations of context effects' Journal of Verbal Learning and Verbal Behaviour 18:645-659
- Tanenhaus, M. K., Dell, G. S. and Carlson, G. 1987 'Context Effects and Lexical Processing: A Connectionist Approach to Modularity' In Garfield, J. L. (ed) *Modularity in Knowledge Representation and Natural-Language Understanding* Cambridge MA/London, MIT Press Pp.83-108
- Tedeschi, P.J. 1981 'Some evidence for a branching-future model' In Tedeschi, P. J. and Zaenen, A. (eds) *Syntax and Semantics 14: Tense and Aspect* New York etc., Academic Press Pp.239-269
- Thomason, R. H. 1970 'Indeterministic time and truth value gaps' Theoria 18:264-281
- Traugott, E. C. 1988 'Pragmatic strengthening and grammaticalization' In Proceedings of the Fourteenth Annual Meeting of the Berkeley Linguistics Society: Parasession on Grammaticalization Pp.406-416
- Traugott, E. C. and Heine, B. (eds) 1991 *Approaches to Grammaticalization* Amsterdam/Philadelphia, John Benjamins
- Traugott, E. C. and König, E. 1991 'The semantics-pragmatics of grammaticalization revisited' In Traugott, E. C. and Heine, B. (eds) *Vol.1: Focus on Theoretical and Methodological Issues* Pp.149-187
- Trueswell, J. C. and Tanenhaus, M. K. 1991 'Tense, temporal context and syntactic ambiguity resolution' Language and Cognitive Processes 6:303-338
- Trueswell, J. C., Tanenhaus, M. K. and Garnsey, S. M. 1994 'Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution' Journal of Memory and Language 33:285-318
- Ulan, R. 1978 'The Nature of Future Tenses' In Greenberg, J. H. (ed) *Universals of Human Language - Vol.3: Word Structure* Stanford, Stanford University Press Pp.83-123. Reprinted from Working Papers on Language Universals (1972) 8:55-100
- Vallduví, E. 1993 *The Informational Component* Philadelphia, University of Pennsylvania Institute for Research in Cognitive Science (IRCS Report 93-98)
- Vallduví, E. 1994 'Detachment in Catalan and information packaging' Journal of Pragmatics 22:573-601
- Van der Auwera, J. 1986 'The possibilities of *may* and *can*' In Kastovsky, D. and Szwedek, A. (eds) *Linguistics Across Historical and Geographical Boundaries. In honour of Jacek Fisiak on the occasion of his fiftieth birthday* Vol. 2 Berlin, Mouton de Gruyter Pp.1067-1076
- Vendler, Z. 1967 *Linguistics in Philosophy* Ithaca N.Y., Cornell University Press

- Wald, B. 1979 'The development of the Swahili object marker: A study in the interaction of syntax and discourse' In Givón, T. (ed) Pp.505-524
- Walton, A. L. 1991 'The semantics and pragmatics of *can*' Linguistische Berichte 135:325-345
- Walz, D. L. and Pollack, J. B. 1985 'Massively parallel parsing: a strongly interactive model of natural language interpretation' Cognitive Science 9:51-74
- Warner, A 1983 'Review of Lightfoot (1979)' Journal of Linguistics 19:187-209
- Warner, A. 1993 *English Auxiliaries: structure and history* Cambridge, C.U.P.
- Wekker, H. Chr. 1976 *The Expression of Future Time in Contemporary British English* Amsterdam, North-Holland
- Wekker, H. Chr. 1980 'Temporal Subordination in English' In Zonnefeld, W. and Weerman, F. (eds) *Linguistics in the Netherlands 1977-1979* Dordrecht, Foris Pp.96-103
- Whiteley, W. H. and Mganga, J. D. 1969 'Focus and entailment: Further problems of transitivity in Swahili' African Language Revue 8:108-125
- Wilson, D and Sperber, D, 1981 'On Grice's Theory of Conversation' In Werth, P. (ed) *Conversation and Discourse* London etc., Croom Helm Pp.155-178
- Wilson, D and Sperber, D, 1986a 'Pragmatics and Modularity' In Papers from the Parasession on Pragmatics and Grammatical theory Pp.67-84 Chicago, Chicago Linguistic Society
- Wilson, D and Sperber, D, 1986b 'Inference and implicature' In Travis, C. (ed) *Meaning and Interpretation* Oxford, Basil Blackwell Pp.45-75
- Wilson, D. and Sperber, D. 1988 'Representation and relevance' In Kempson, R. M. (ed) Pp.133-153
- Wilson, D and Sperber, D, 1993a 'Linguistic form and relevance' Lingua 90:1-25
- Wilson, D and Sperber, D, 1993b 'Pragmatics and time' UCL Working Papers in Linguistics 5:277-298
- Wilson, P. M. 1970 *Simplified Swahili* Nairobi etc., East African Literature Bureau