A relevance theory perspective on grammaticalization

STEVE NICOLLE*

Abstract

Sperber and Wilson's relevance theory draws a distinction between conceptual and procedural encoding. This is a cognitive distinction, according to which conceptual encoding contributes to the construction of conceptual representations and procedural encoding constrains the inferential computations performed over these. Recent work within the relevance theory framework (e.g., Haegeman 1989, Klinge 1993, Nicolle 1997a) has characterized grammatical markers of modality, tense, and aspect as exponents of procedural encoding. Although these accounts are synchronic, the general claim that grammatical markers encode procedural information should also be compatible with evidence concerning the historical development of grammatical markers (a process known as grammaticalization). In this article, I demonstrate that a procedural characterization of grammatical markers is not only compatible with research into grammaticalization, but also sheds light on the following problematic issues. First is the question of whether grammaticalization is semantically gradual (as is generally assumed) or semantically instantaneous as Givón (1991) claims; I provide evidence in favor of Givón's view. Secondly, a procedural analysis suggests an answer to the question of what semantic mechanism initiates grammaticalization. Finally I demonstrate how the distinction between conceptual and procedural encoding accounts for the phenomenon of semantic retention (residual lexical meaning) in grammatical markers.

Keywords: relevance theory; grammaticalization; language change.

1. Conceptual and procedural encoding in relevance theory

This article investigates grammaticalization (the process whereby grammatical expressions such as tense, aspect, and modality markers develop out of lexical expressions) from the perspective of the relevance theory

distinction between conceptual and procedural encoding. Relevance theory (Sperber and Wilson 1986, 1995) provides a cognitive perspective on communication and on utterance interpretation in particular, and the conceptual/procedural distinction is motivated by some general assumptions underlying cognitive science.

Most, if not all, current theories of cognition recognize 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. A further assumption is that humans are rational information processors, where 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), thereby increasing the chances of successful processing. As a result, many cognitive theories (such as mental models [Johnson-Laird 1983], Fodor's [1983] modularity hypothesis, and relevance theory) assume that human information processing 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.[1]

Utterance interpretation is a particular form of inferential 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 (in the sense described above), 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 such a way as to reduce to a minimum the effort required to process them. This is the motivation for the conceptual/procedural distinction in relevance theory.

Wilson and Sperber (1993: 2) describe the motivation for the distinction between conceptual and procedural encoding as follows:

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.

The construction and manipulation of conceptual representations which Wilson and Sperber (1993) mention derives from linguistically encoded stimuli and a set of assumptions (constituting a context). Conceptual information gives rise to conceptual representations which provide the input to inferential processes of utterance interpretation, whilst procedural information provides constraints on the inferential computations performed over these conceptual representations. All linguistically encoded information is viewed as being of one of these two general types; there is no intermediate information type.

Procedural information was first posited by Blakemore (1987, 1988) who suggested that the purpose of discourse connectives such as *so*, *after all*, and *but* is to constrain the inferential processing of propositions. The problem with such expressions for traditional truth-conditional theories is that, despite being linguistically encoded, they do not contribute to the truth conditions of utterances containing them. By introducing the notion of procedural encoding, Blakemore could characterize *so*, *after all*, and *but* as constraining the derivation of implicatures (inferred conclusions) from utterances containing them. In constraining the inferential processing of propositions in this way, the procedural information encoded by *so*, *after all*, etc. has the effect of reducing the computational effort required of an addressee to derive implicatures, thereby increasing the chances of successful utterance interpretation.

Relevance theory recognizes that inference is involved not only in the processing of propositions to derive implicatures, but also in the construction or recovery of propositions in the first place. Wilson and Sperber (1981) demonstrated that propositions are not recovered simply through linguistic decoding, but that inferential processing is crucial to the resolution of vagueness and the recovery of ellipsed material, as well as to reference assignment and disambiguation, all of which contribute to the identification of any proposition expressed (see also Carston 1988, 1993, Recanati 1989). It is therefore to be expected that procedural encoding will constrain not only the inferential processing of fully propositional conceptual representations (as Blakemore claims discourse connectives do), but also the inferential computations performed over subpropositional conceptual representations which result in the construction of propositions.

Wilson and Sperber (1993) proposed that personal pronouns might perform such a function by constraining the largely inferential process of reference assignment, and a similar account is proposed for referring expressions in general by Ariel who argued that "referring expressions are no more than guidelines for retrievals" (1988: 68). In some recent relevance theory work (Haegeman 1989, Klinge 1993, Nicolle 1997a) this approach has been extended to accounts of modality, tense, and aspect (MTA) markers. Although there are considerable differences between these accounts, they all characterize MTA markers as exponents of procedural encoding.

This common procedural characterization of both pronouns and modality, tense, and aspect markers reflects the syntactic and semantic

parallels drawn in the literature between pronominal and temporal anaphora (Carston 1988, Partee 1973, 1984) and between pronominal anaphora and modality (Roberts 1989). Both pronominal and temporal reference are determined largely inferentially, as is the existential status of situations and events (that is, whether they are to be represented as actual, verified or unverified, hypothetical, conditional, etc.). Just as pronominal expressions merely guide addressees towards their intended referents, so too do tenses underdetermine the temporal reference of situations and events. In both cases, the role of the indexical expressions is to constrain the inferential process of reference assignment rather than to determine this.

In English, this can be illustrated as follows. Even when a tense/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 example, the temporal reference of the events in the following examples is determined largely as a result of pragmatic considerations, rather than by the "present perfect" alone; the event referred to in (1) is assumed to be recent (say, within a few hours of the time of the utterance) whereas the event referred to in (2) may have occurred many years before the utterance.

(1)   Mary has eaten.
(2)   Mary has climbed Everest.

The importance of inference in the determination of temporal reference is even clearer in languages such as Chinese, which does not have formally grammaticalized tense markers. 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 (cf. Li and Thompson 1979), is pragmatically determined. Tense markers, in those languages which have them, may thus be characterized as merely imposing constraints on the determination of temporal reference. Similarly, markers of modality may be viewed as encoding constraints on the existential status of situations and events. Conversely, it is difficult to see how markers of tense and modality could be characterized conceptually. Take, for instance, examples (1) and (2); the "present perfect" does not encode information about itself so much as about the interpretation of the events described by (1) and (2), say, that the events [Mary eat] and [Mary climb Everest] are meant to be represented as occurring at some time in the past whilst having present relevance. As a result of these considerations, grammatical markers of tense and modality may be characterized as exponents of procedural encoding, constraining the inferential processing of conceptual representations of situations and events.

In summary, relevance theory recognizes two distinct linguistically encoded information types. Conceptual encoding gives rise to conceptual representations which provide the input for inferential processing, and procedural encoding provides constraints on the inferential processing of conceptual representations. Some procedural expressions (for example *so*, *after all*, and *but*) help addressees recover implicatures by constraining the inferential computations performed over propositional conceptual representations. Other procedural expressions, including pronouns and modality, tense, and aspect markers, constrain the inferential computations performed over subpropositional conceptual representations, thereby guiding addressees toward the identification of propositional forms.

Finally, I will mention two complementary aspects of the conceptual/procedural distinction which are particularly relevant to grammaticalization. First, conceptual and procedural information constitute two distinct information types; there is no information type intermediate between them. Secondly, following Groefsema's (1992: 220) argument that both conceptual and procedural information should be represented in the language of thought, Nicolle (1997b) has proposed that there is in principle nothing to stop a single expression encoding both conceptual and procedural information, since both stem from the same set of ontological categories (in the sense of Jackendoff [1983], illustrated in section 4.1).[2] (In section 3, I will argue that this second aspect resolves problems for a procedural account of grammaticalization that the first causes.)

2.   Grammaticalization and procedural encoding

Grammaticalization (or "grammaticization"), 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). Recently, the role of constructions as well as isolated lexical items has been emphasized (e.g., Lehmann 1995 [1982]: viii, Bybee et al. 1994: 4),[3] and this paper reflects this trend, for example in discussing the *be going to* construction and the development of a new aspectual marker in Kiswahili from an aspect + verb construction.

Grammaticalization results in semantic, syntactic, morphological and phonological changes in the grammaticalizing expression (the "gram"). Semantically, the developing gram moves from encoding a specific semantic content, appropriate to a restricted range of contexts, to encoding a very generalized, 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 1975). 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 grammaticalization leads to increased rigidity in syntactic position and scope relations, 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 (e.g., *you'll*) and affixed to the reduced allomorph of *not* (itself a gram), as *won't*. Affixation to a lexical item is not a prerequisite for phonetic reduction as *gonna*—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 grammaticalization to be discussed here is the process whereby modality, tense, and aspect (MTA) markers develop from lexical verbal constructions. MTA markers are all grams, and so, from the perspective of the relevance theory conceptual/procedural distinction, grammaticalization involves a shift from conceptual encoding to procedural encoding in a single expression over time. This is not to say that all exponents of procedural encoding are grams; discourse connectives such as *so* and *after all*, which Blakemore (1987, 1988) characterizes as encoding procedural information, are not grammatical markers. *So* and *after all* constrain the processing of fully propositional conceptual representations with respect to the implicatures which a speaker intends an addressee to compute, whereas grams, such as modality, tense, aspect, and case markers, constrain the construction or identification of propositional conceptual representations.[4]

The relevance theory characterization of procedural encoding must be reconciled with the following aspects of grammaticalization. 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, formally, grammaticalization is a gradual process with many individual expressions occupying intermediate positions on a continuum between fully lexical and fully grammaticalized.

Secondly, procedural encoding has been characterized as constraining the inferential processes essential to utterance interpretation. We would therefore expect to see a link between the onset of grammaticalization and the inferential processes involved in utterance interpretation; I shall therefore investigate the causes of grammaticalization.

Finally, I will take into consideration the fact of semantic retention, whereby conceptual information encoded by the lexical source expressions of certain grams is still accessible in the use of those grams in certain contexts. I shall address each consideration in turn with particular reference to the English *be going to* future construction and various Kiswahili constructions.

3.   The gradualness of grammaticalization

It is generally accepted within the literature that grammaticalization is a gradual process (see for example Lichtenberk 1991: 37, Hopper and 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 grammaticalized and grammaticalizing expressions are distributed, clustering at certain points (as auxiliaries or affixes for example). It is further assumed that the various changes involved in grammaticalization—both formal and semantic—occur in parallel:

> both types of formal change in grammaticalization parallel the main types of semantic change in grammaticalization. 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)

This characterization of grammaticalization as both formally and semantically gradual 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, then there must be an intermediate position between conceptual and procedural encoding. In response to this problem, I shall examine evidence which contradicts the underlying assumption that grammaticalization is semantically gradual.

3.1.   Counter-evidence from verb serialization

The underlying assumption that semantic development occurs in parallel with the formal aspects of grammaticalization (Bybee et al. 1994: 106)

is challenged in Givón (1991). Givón looked at verb serialization, 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 serialization can function not only as co-lexicalization (for example, the translation of English *fetch* as *aller chercher* in French) but grammatically, for example by marking tense and aspect, as in (3),

(3) a. he stay work   (durative)
         'he is working'
     b. he go work   (future)
         'he will work'
     c. he work finish   (perfective)
         'he has already worked'
         [Givón 1991: 83]

and by assigning nominal case-roles (in the following examples [4a–d] it is the verb immediately preceding the object NP which grammaticalizes, regardless of the order of the two verbs).

(4) a. she take-stick break   (patient)
         'she broke the stick'
     b. she walk go-market   (locative)
         'she walked to the market'
     c. he work give-her   (benefactive)
         'he worked for her'
     d. she take-knife cut meat   (instrumental)
         'she cut the meat with the knife'
         [Givón 1991: 82]

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 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 (cf. Bradshaw 1982, Crowley 1987). 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 (1991: 116) are summarized as follows:

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.

What these findings demonstrate is that lexical expressions can be used with the same functional load as grammaticalized 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)

This conclusion is too strong, since Givón's findings prove only that grammaticalization *need* not be a gradual process; it is possible that only serial-verb constructions lead to instantaneous grammaticalization.[5] Given that at least some cases of grammaticalization are instantaneous, Givón's findings raise the following question. When a lexical expression is used, as Givón puts it, in a frame that "intends" it as a grammatical marker, and is interpreted as such, is the lexical meaning of the expression still recovered by the addressee, and if not, why not?[6] 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 *recognized* or *recovered*.

By analogy, I suggest that, in the case of a formally lexical expression used functionally/semantically as a gram, the newly encoded procedural information is automatically recovered each time the expression is processed, since it provides an effort-reducing 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 criterion of consistency with the principle of relevance (since recovering and inferentially enriching the conceptual information also encoded would increase processing effort and thereby reduce 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 and Sperber 1988: 141)

Givón (1991) goes on to suggest that the supposedly gradual nature of grammaticalization 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 grammaticalization, 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, there are 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 grammaticalization mentioned in section 2 (syntactic rigidity, phonetic reduction, etc.).

3.2. *Evidence from Kiswahili*

Further evidence that the semantic changes involved in grammaticalization do not necessarily occur in parallel with the gradual formal changes that accompany it is provided by the development of the Kiswahili aspectual marker *-mesha-*.

MTA markers in Kiswahili are encoded as preverbal prefixes:

(5) *Ni+   me+   soma*
     SP   PERF read
     'I have 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 development (mentioned) 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).[7] This is a recent development in Kiswahili; *-mesha-* is mentioned in the grammars of neither Ashton (1947) nor Polomé (1967).

The following examples illustrate in turn the use of *me+kwisha* (*i*) in a serial-verb construction either with or without functional (semantic) grammaticalization (6), (*ii*) as a fully grammaticalized prefix with the form *-mekwisha-* (7), and (*iii*) as a reduced prefix with the form *-mesha-* (8); each of these constructions is current in contemporary Kiswahili.

(6) *A+   me+   kwisha ku+ soma*
     SP   PERF kwisha INF read
     'S/he has finished reading'; 's/he has already read.'

(7) *A+   mekwisha+   soma*
     SP   mekwisha   read
     'S/he has already read.'

(8) *A+   mesha+   soma*
     SP   mesha   read
     'S/he has already read.'

The construction in (7) with *-mekwisha-* treated as an MTA prefix is sometimes treated as a variant of the construction in (6) analyzed as the auxiliary verb *kwisha* with the MTA prefix *-me-* but followed by a bare verb stem (i.e., minus the nonfinite *ku-* prefix). The following transcription from Maw and Kelly (1975: 110–111) illustrates this practice:

(9) *Je,   ndio   wakati huu u+   me+   kwisha timiza mihadi*
     INT   really time   that SP   PERF kwisha fulfill promise
     yenu,   au u+   me+   kwisha vunja mihadi   yenu?
     your (pl) 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 (7) is best analyzed as an 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 (6) 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):

(10) *Ni+ me+ kwísha*
      SP   PERF finish
      'I have finished.'

The fact that the stressed syllable /kwí/ has been lost in the reduced form of the grammatical marker *-mesha-* suggests that, when used grammatically, serial-verb constructions such as (6) (repeated here as [11] with stress indicated) function prosodically as a single verbal group with stress on the penultimate syllable of the main verb only.

(11) *A+ me+ kwisha ku+ sóma*
      SP PERF kwisha INF read
      'S/he has already read.'

In their transcription of (9), Maw and Kelly (1975) observed that the syllable /sha/ functioned as what they term a "salient syllable". Maw and 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 and Kelly 1975: 7). This entails that the syllable /sha/ took secondary stress in the utterance of (9) rather than /kwi/ since /sha/ marked the beginning of a salient.

Once formal grammaticalization has taken place and *-mekwisha-* is realized 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 marker (OM):

(12) *U+ mekwisha+ mw+ ona boy wa   somo yako?*
      SP mekwisha OM   see boy from class your
      'Have you seen the boy from your class yet?'
      [*Rosa Mistika (RM)*, p. 34; see Key to Kiswahili sources]

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-* (13), *-ki-* (14), *-me-* (15), *-mekwisha-* (16) (the *-mekwisha-* prefix being more common in the main clause of complex VPs than *-me-*), and the negative forms *ha-ja-* (17) and *ha-i* (18), these latter two consisting of an initial prefix *ha-* plus a pre-stem prefix *-ja-* and a verbal suffix *-i* respectively.[8] The subject of the auxiliary is the same as that of the main verb (in later examples *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 *Rosa Mistika* [*RM*], p. 6).

(13) *Tangu Ijumaa mpaka Jumapili Zakaria a+ li+   kuwa a+*
      from   Friday until Sunday Zakaria SP PAST be   SP
      na+   kwenda kunywa ...
      PROG go   drinking
      'From Friday to Sunday Zakaria went drinking ...'

(14) *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+
      Zakaria SP PAST be   SP PROG dance SP PAST
      kuwa a+ ki+   rukaruka.
      be   SP PROG jump-jump
      'Whilst the children were singing, Zakaria was dancing—he was jumping up and down.'

(15) *Zakaria a+ li+   kuwa a+ me+   shindwa hata kujenga*
      Zakaria SP PAST be   SP PERF defeated even to-build
      nyumba ya maana.
      house   of import
      'Zakaria was incapable even of building a decent house.'

(16) *Kwa wakati huo wawili wa+ li+   kuwa wa+ mekwisha+*
      By   time   that two   SP   PAST be   SP   mekwisha
      uzwa.
      be-sold
      'By that time two (of them) had already been sold.'

(17) *a+ li+   kuwa ha+ ja+   toa hata chapa*
      SP PAST be   SP NEG 'do' even stroke
      ku+ wa+ lipia   watoto   ada ya shule.
      to   OP   pay-for children school fees

'he had not done even a stroke (of work) to pay for his children's school fees.'

(18) *Stella a+ li+ kuwa ha+ wez+ i ku+ ficha siri ...*
      Stella SP PAST be -SP be able NEG to keep secret
      'Stella was unable to keep a secret ...'

The aspectual function of *-mekwisha-* in conjunction with a tensed form of *kuwa* is also possible when *kwisha* is morphologically an auxiliary verb with the perfect marker *-me-*:

(19) *a+ li+ kuwa a+ me+ kwisha ku+ agiza*
      SP PAST be SP PERF kwisha INF instruct
      'he had already given instructions'
      [*Mzimu wa Watu wa Kale (MWK)*, p. 1]

Finally, as evidence of the total integration of *-me(kwi)sha-* into the MTA paradigm, it can occur as an aspectual prefix in a complex VP, with a following OM:

(20) *wasichana wa+ li+ kuwa wa+ mesha+ m+ shangilia*
      girls SP PAST be SP mesha OM congratulate
      sana Rosa.
      much Rosa
      'the girls had already often congratulated Rosa.'
      [*RM* p. 35]

What these examples illustrate is that three structurally distinct forms current in contemporary Kiswahili (*-mekwisha+*infinitive, *-mekwisha-*, and *-mesha-*), can all perform the same semantic function of aspect marking. The fact that *-mekwisha-* and *-mesha-* both behave syntactically as typical MTA prefixes (which until recently was not the case) indicates that grammaticalization has occurred, and the fact that the stressed syllable in the lexical verb *kwisha* has been lost supports Givón's iconicity principle for serial-verb constructions (that is, when used with the same functional load as a grammatical marker, /kwi/ did not take primary stress).

3.3. *Summary*

Givón's characterization of grammaticalization as formally gradual but semantically instantaneous is compatible with the claim that grammaticalization involves a shift from conceptual encoding to procedural encoding within a single expression. A gram part way along the (formal) grammaticalization cline cannot be said to be semantically intermediate between lexical and grammatical if there is no information type intermediate between the conceptual and the procedural. To characterize

an expression as semantically intermediate in this way would also contradict the evidence from Givón (1991) and from the development of the *-mesha-* prefix in Kiswahili that grammaticalization can be instantaneous. Neither is it necessary to characterize a gram, or an expression used grammatically, as encoding exclusively either conceptual (lexical) information or procedural (grammatical) information; to do so would result in polysemy every time that, as Givón (1991: 123) puts it, "a lexical item is used in a frame that intends it as a grammatical marker". If an expression can (as Nicolle 1997b proposed) encode both conceptual information and procedural information, then grammaticalization (the development of a procedural semantics) need not result in the loss of lexical (conceptual) semantic features.

This still leaves two questions: first, how and why does an expression come to encode procedural information in addition to its conceptual semantic content, and secondly, does formal grammaticalization, if and when it occurs, invariably lead to a loss of conceptual semantic content? In the following sections I discuss first the causes of grammaticalization, and secondly semantic retention.

4. The causes of grammaticalization

The cause of any particular case of grammaticalization is likely to be complex, with language users (both speakers and hearers), semantic and pragmatic factors, structural changes, and possibly influence from other languages all playing a role.[9] What this section is concerned with, therefore, is the motivations and mechanisms which are necessary for grammaticalization to occur; it is not assumed that any one factor is a sufficient cause of grammaticalization in itself.

So far, I have assumed that grammaticalization is motivated by semantic rather than by formal changes. Although this is a widely held position in the literature (cf. Bybee et al. 1994: 281–301), advocates of the autonomous syntax approach (e.g., Lightfoot 1991, Warner 1993) suggest that semantic developments do not have "clear-cut independence and priority" (Warner 1993: 196) from formal, morphosyntactic changes. Concerning the development of the English modal auxiliaries, Warner (1993: 197) states, "grammaticalisation ... is not here a semantically led development". Warner's claim is that independent formal changes to a group of lexical expressions (the premodals) paved the way for semantic change. However, he also says that grammaticalization "involves general principles of class formation (internal coherence and external opposition) which would presumably also apply to 'non-grammatical' classes" (1993: 196). Formal changes resulting in grammaticalization cannot therefore be

the sole mechanism underlying grammaticalization, otherwise any case of morphosyntactic class formation could result in grammaticalization; rather, Warner is claiming that semantic developments, while crucial, cannot result in grammaticalization without accompanying formal changes such as class formation. Lightfoot (1991: 148) also stresses the importance of formal changes in the grammaticalization of the English modal auxiliaries, claiming that meaning change in the premodals

> was entirely a by-product of the [syntactic] recategorization: once *shall, can,* etc. were classed as INFL, they could not occur with direct objects, and consequently the meanings they had with direct objects were automatically lost.

This explains the loss of meaning: *sculan* (the source of *shall*) had the lexical meaning 'to have to pay (something)', and *cunnan* (the source of *can*) meant 'to know (something)'; it also provides an account of how the development of a new morphosyntactic class could have acted as a catalyst for the grammaticalization of the members of this class. However, an account of the loss of lexical meaning in the premodals fails to explain the development of grammaticalized meaning; as Langacker (1990: 16) notes, "as an element becomes grammaticized ... it undergoes a change of meaning rather than becoming meaningless". The autonomous syntax position also fails to explain why grammaticalization affects similar semantic classes in similar ways in quite unrelated languages (Hopper and Traugott 1993: 68).

Further evidence in favor of the view that semantic change drives formal change, rather than being driven by it, comes 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 fully grammaticalized expressions with full lexical semantics. The assumption that semantic change drives formal change is also supported by Givón's (1991) observations that formal features of grammaticalization develop after functional changes rather than in parallel with them.

In terms of the conceptual/procedural distinction, I will argue that the semantic change driving grammaticalization (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. In other words a grammaticalizing expression comes to constrain the interpretation of the associated utterance in some way, whilst continuing to encode information with the potential to give rise to a conceptual representation. The question to be addressed in this section is what causes lexical expressions encoding conceptual information to come to encode procedural information as well.

Bybee, Perkins, and Pagliuca (1994: 282) identify five mechanisms of semantic change involved in the development of grams marking modality, tense, and aspect—some characteristic of early stages of the grammaticalization process, others of later stages. Assuming that the shift from conceptual to procedural encoding is the driving force behind the formal changes involved in grammaticalization, I shall concentrate on describing those mechanisms operative in the early stages of grammaticalization 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 conventionalization of implicature, "generalization", "harmony", and "absorption" (of features of the linguistic contexts in which a gram prototypically occurs).[10] I shall briefly describe each before looking in more detail at those involved in the early stages of grammaticalization.

4.1.   *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, motivated by analogy and/or iconic relationships between the categories. These can be viewed as primitive ontological categories similar to those proposed by Jackendoff (1983). Heine, Claudi, and Hünnemeyer (1991: 161) illustrate the effect of metaphorical extension on the use of the Ewe lexeme *megbé* 'back'; in example (21) *megbé* denotes a body part (OBJECT), in (22) a location (SPACE), either as an adverb (22a) or as a postposition (22b), in (23) a temporal relation (TIME), and finally in (24) it means 'mentally retarded' (QUALITY).

(21)   *é-pé   megbé fá*   OBJECT
        3SG-POSS back   be-cold
        'His back is cold.'

(22) a. *é   le xɔ   á   megbé*   SPACE
        3SG is house DEF behind
        'He is at the back of the house.'

     b. *é   nɔ megbé*   SPACE
        3SG stay behind
        'He stays back.'

(23)   *é   kú le é-megbé*   TIME
        3SG die be 3SG-behind
        'He died after him.'

(24)   *é   tsi   megbé*   QUALITY
        3SG remain behind
        'He is backward/mentally retarded.'

Example (24) 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, reflects some overlap between consecutive semantic categories.

Despite the attention paid to metaphorical extension as a potential cause of grammaticalization, Bybee, Perkins, and Pagliuca (1994: 283) observed that very few examples of grammaticalization of modality, tense, or aspect in their study could be accommodated in the metaphorical model, and of those that could, the change occurred very early in the grammaticalization 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, grammaticalization involves a special kind of metaphorical extension: metonymy, or the conventionalization of implicature. (See also Hopper and Traugott 1993: 77–87.) Carey, who describes the role of the conventionalization of implicature in the early development of the English perfect, characterizes this mechanism as follows:

When semantic change arises from the conventionalization 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)

In relevance theory, inference is recognized as playing an important role not only in the identification of (conversational) implicatures but also in the identification of any proposition(s) expressed (termed "explicatures"). That is, invited inferences contribute both to explicatures—which are developments of the logical forms (or semantic representations) encoded by utterances—and to implicatures, which are developments of explicatures on the basis of contextual assumptions and the presumption of relevance. I shall therefore follow Bybee, Perkins, and Pagliuca (1994) in referring to what has been called the conventionalization of implicature, but which is in fact the conventionalization of invited inferences more generally, simply as "inference".

4.2. Generalization, harmony, and absorption

Three further mechanisms of semantic change are discussed in Bybee, Perkins, and Pagliuca (1994), the first of which is generalization, "the loss of specific features of meaning with the consequent expansion of appropriate contexts of use for a gram" (Bybee et al. 1994: 289). Generalization can be described in terms of the loss of one feature of meaning; this is illustrated in the development of can from being a marker of purely mental ability to general ability and finally to representing root possibility

(Bybee et al. 1994: 290):

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

Generalization 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 generalized 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 grammaticalization (Bybee et al. 1994: 294) and results in semantic redundancy. Absorption of features of the linguistic contexts in which a gram prototypically occurs is also a feature of the later stages of grammaticalization. Absorption results in a shift in the function of an already existing gram; for example, Schwenter (1994) reports on the development of a Spanish marker of anteriority into a perfect, which he attributes in part to the absorption of the temporal context in which the grammaticalizing expression normally occurs.

Both harmony and absorption are involved in the final stages of formal grammaticalization in which a gram can come to be used so frequently, including when it harmonizes 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 a "zero gram", which absorbs contextual meaning pertaining to the domain constrained by the overt gram. Since harmony and the absorption of contextual meaning are only involved in the later stages of grammaticalization they will not be discussed further. Instead, I will focus on metaphorical extension, "inference", and generalization, and attempt to determine which of these is/are involved in the shift from purely conceptual encoding to procedural encoding and hence grammaticalization.

4.3. The mechanism of semantic change in grammaticalization

Heine, Claudi, and Hünnemeyer (1991) claim that metaphorical extension is one of the main, if not the only, mechanisms of semantic change

involved in grammaticalization. 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 *megbé* discussed above), are arguably not instances of grammaticalization (cf. Bybee et al. 1994: 284). At best, metaphorical extension often precedes subsequent grammaticalization and may be a prerequisite for it. Secondly, the clearest cases of grammaticalization 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, Perkins, and Pagliuca (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 (25) and the probability reading in past or present contexts such as (26).

(25)   The letter must arrive sometime next week.
(26)   The letter must have been in the mail.

Since—they reason—an epistemic interpretation is not available in (25) and a deontic interpretation is not possible in (26), 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 (25) can receive an epistemic interpretation:

(25)′ 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 (25)′ 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 (26) 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):

(26)′ 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, Perkins, and Pagliuca (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) characterizes it as crucially involving "explicit coding of relevance and informativeness that earlier was only covertly implied". In other words, grammaticalization involves the conventionalization of invited inferences involved in uncovering covert (i.e., implied) meaning, even in the absence of direct contextual support. Thus, for example, from an utterance of (25)′ 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 grammaticalization, the inference to an epistemic interpretation can be made even in the absence of a deontic statement from which epistemic probability could be inferred.

Procedural encoding constrains the inferential processes involved in utterance interpretation in just this way, suggesting that "inference" (the conventionalization of invited inferences) is one mechanism of semantic change which results in a shift from purely conceptual encoding to procedural encoding also. "Inference" can be illustrated in greater detail with respect to the development of the modal auxiliaries *will* and *shall* and the future marker *be going to*.

The development of *will* and *shall* from lexical verbs (encoding conceptual information) to grams (encoding procedural information) and then their further development as grams can be described in terms of the following stages (the relevant mechanism of semantic change at each stage is indicated in parenthesis):

i.   Lexical: obligation (*shall*), desire (*will*) used with first person subjects.
                     (Metaphorical extension results in ...)
ii.   Lexical: intention with first person subjects (*will* and *shall*).
                     ("Inference" results in ...)
iii. Grammatical: prediction about future events with first person subjects.
                     (Generalization results in ...)

iv. Grammatical: prediction about future events (any subject).
     (Further generalization results in ...)
v. Grammatical: prediction in general (*will* only).

Here, generalization is involved in the later stages of grammaticalization process, after the shift to procedural (grammatical) encoding has taken place. Generalization also seems to be involved in the early semantic development of the *be going to* construction from the imperfective form of the lexical movement verb *go* (encoding conceptual information), plus allative *to*, to a grammatical marker of future time reference relative to some temporal reference point (encoding procedural information):

i. Lexical: physical progression. The subject is in the process of progressing towards a physical goal.

(Generalization results in ...)

ii. Lexical: general progression. The subject is in the process of progressing towards a goal.

("Inference" results in ...)

iii. Grammatical: future time reference. The subject will, at some future time, achieve a goal.

The goal towards which the subject is progressing is a physical location in (i) and the realization of a further situation or event in (ii). The change from progressing physically towards a goal to progressing in a non-physical sense towards a goal, both of which are lexical rather than grammatical meanings, can be viewed as due to generalization (the loss of the physical component of meaning). If the subject is in the process of progressing towards the realization of a further situation or event, the subsequent realization of that situation or event can be inferred. The grammaticalization of *be going to* as a marker of future time reference can be achieved through the conventionalization of this invited inference.[11] It is noticeable that it is the *be going to* construction which has undergone grammaticalization, rather than the lexical item *go* in isolation (see section 5.1. for further discussion). This further reflects the role of "inference" in grammaticalization, since "inference" derives from the interpretation of phrases in context, in contrast to metaphor which operates on lexical items independently of the contexts in which they occur.

I have found no case of generalization changing a lexical to a grammatical expression (i.e., initiating procedural encoding), suggesting that generalization as a mechanism of semantic change is not itself the mechanism by which grammaticalization is activated through the shift from

conceptual 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"—the conventionalization of invited inferences—as the one sure mechanism of semantic change driving the shift to procedural encoding which initiates grammaticalization. Of course, there are many invited inferences generated by utterances of many different linguistic expressions, so the question still remains of what causes only some of these to become conventionalized with resultant grammaticalization. A motivating factor determining what types of inferences are likely to result in grammaticalization may be "subjectification", that is, the tendency of grammaticalized meanings to become increasingly situated in the speaker's subjective belief/attitude toward the situation being spoken about (Traugott 1988: 410). As an explanation of why only certain types of inference typically result in grammaticalization, subjectification can be viewed as complementary (rather than contradictory) to the characterization proposed here of grammaticalization in terms of a shift from conceptual to procedural encoding, resulting from the conventionalization of invited inferences. The role of subjectification in grammaticalization is discussed in detail in Langacker (1990) (see also Sanders and Spooren 1996) and so will not be discussed further here.

5. Semantic retention

I have so far argued that grammaticalization is semantically instantaneous but formally gradual, and that the mechanism of semantic change which triggers grammaticalization is the conventionalization of invited inferences. I have suggested that any procedural information encoded by an expression is automatically recovered (in addition to being merely activated on decoding) 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 (in accordance with the criterion of consistency with the principle of relevance) so as to minimize processing effort. However, when a formerly (or formally) 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).

Some grams, such as the English past tense suffix, *-ed*, encode no lexical content at all, suggesting that, as a gram develops over time, the conceptual information it initially encodes may become increasingly inaccessible, until such time as it disappears altogether and the gram encodes only

procedural information.[12] (This process may be initiated as a result of formal changes to the grammaticalizing expression such as the loss of lexical meaning in the English premodals as described by Lightfoot [1991] above.) However, the relevance theory account predicts that from the moment an expression first encodes procedural information until the point where, like the -*ed* past tense marker, it no longer encodes any conceptual content, both conceptual and procedural information should be encoded and recoverable. If the model of grammaticalization outlined here is correct, we should expect to find grams exhibiting lexical semantic characteristics in certain utterances, in addition to performing their grammatical functions. This is in fact what we find in cases of "semantic retention" (Bybee and Pagliuca 1987), whereby "certain more specific semantic nuances of the source construction can be retained in certain contexts long after grammaticalization has begun" (Bybee et al. 1994: 16).

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 within the relevance theory framework. 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 and Pagliuca (1987) and Bybee, (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 account of *be going to* proposed below.)

In the remainder of this section, I will discuss semantic retention in the *be going to* future construction. After providing evidence for treating *be going to* as a grammatical marker, I will discuss interpretations of utterances containing *be going to* which provide evidence for semantic retention, before concluding with a relevance theory account of why these interpretations arise in some contexts but not in others.

5.1. *Arguments for treating* be going to *as a grammatical marker*

The *be going to* construction exhibits many of the typical characteristics, both formal and semantic, of grammatical markers. Formally, it has a

phonetically reduced allomorph *gonna*, and exhibits a high degree of syntactic rigidity, in that it only takes bare, nonfinite verbal complements. Both characteristics are illustrated in (27):

(27) I'm gonna see a film at the cinema.

In contrast, the nongrammaticalized present imperfective form of *go* with locative *to* takes nominal complements and has no phonetically reduced allomorph:

(28) a. I'm going to the cinema to see a film.
      b. *I'm gonna the cinema to see a film.[13]

Semantically, *be going to* encodes information to treat the situation or event described by an utterance in which it occurs as future relative to some temporal reference point (a typical semantic field for grams). The temporal reference point may be the moment of speaking or some other time, either in the past or the future:

(29) a. I'm going to visit Mary.
      b. I was going to visit Mary two hours ago.
      c. I am going to visit Mary after I finish work.

In (29a) the event [I visit Mary] is envisaged subsequent to the moment of speaking. In (29b) the event [I visit Mary] is envisaged as subsequent to some time in the past; the temporal adverbial *two hours ago* specifies either the temporal referent point or the time at which the event [I visit Mary] was envisaged as occurring. Example (29c) envisages the event [I visit Mary] occurring subsequent to the time at which the speaker finishes work. The *be going to* construction therefore consists of two components: a form of auxiliary *be* providing information relating to the temporal reference point, and the *going to* component which indicates future time reference relative to this reference point.

The characterization of *be going to* as a (relative) future marker is supported by the following considerations. First, it has been observed (Bybee et al. 1991: 18, Bybee et al. 1994: 244, Fries 1927, Ultan 1978) that, cross-linguistically, movement verbs (such as *go*) are a common source of future grams. Secondly, Bybee, Perkins, and Pagliuca (1994: 11) observe that future markers 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 *to*).[14] Since movement verbs are also a common source of pasts and progressives, Bybee, Pagliuca, and Perkins (1991: 30) suggest that movement alone is not sufficient to give rise to future markers; imperfectivity and an allative component are also

required. *Be going to* has all three, and it is the combination of these features which has resulted in the development of *be going to* as a grammatical marker of (relative) future time reference.[15]

5.2.   *Semantic retention in* be going to

In certain contexts, *be going to* gives rise to some typically lexical interpretations which are not derivable from enrichment of its procedural semantic content but which are not implicatures (purely inferred interpretations) either. Two examples of these interpretations are the overtones of prior intention and inevitability which utterances containing *be going to* may convey.[16]

Intention originating prior to a request is suggested if an affirmative response to the request contains *be going to*, in contrast to a response containing *will*, as the following minimal pair demonstrates.

(30)   Can somebody visit John tomorrow?
   a.   I'm going to visit him.
   b.   I'll visit him.

Example (30a) suggests that the speaker was already intending to visit John at the time of the request, whereas (30b) suggests that the speaker had not been intending to visit John but decides to on being asked.

The prior intention interpretation of *be going to* also accounts for the unacceptability of (31a) as it is impossible to intend to answer a telephone before it rings; (31b) with *will* is acceptable, however:

(31)   [Immediately following ringing of telephone:]
   a. ?I'm going to get it.
   b.   I'll get it.

Although utterances with *be going to* often suggest prior intention and utterances with *will* often suggest intention originating subsequent to a request, only the prior intention interpretation with *be going to* is encoded. This is demonstrated by the fact that the prior intention interpretation derived from utterances containing *be going to* is not cancellable, while the subsequent intention interpretation derived from utterances containing *will* is. Recanati (1989) notes that cancellability, one of the six characteristics of Gricean conversational implicature, is not wholly superfluous within the relevance theory framework; although it does not distinguish implicatures from explicatures, it does at least indicate the presence of pragmatically determined aspects of utterance meaning (that is, purely inferred interpretations). Thus, if an interpretation is

purely inferred, it should be cancellable, but if an interpretation results from decoding a linguistic expression it will not be cancellable. The overtone of prior intention conveyed by utterances containing *be going to* as in (32a) cannot be cancelled, indicating that this is at least partially encoded, whereas the subsequent intention interpretation derived from utterances containing *will* such as (32b) is cancellable and is therefore inferred:

(32)   Can somebody visit John tomorrow?
   a.   ?I'm going to visit him, but I wasn't intending to.
   b.   I'll visit him, in fact I was already intending to.

In addition, an overtone of inevitability can arise with *be going to* but not with *will*. In the following minimal pair, (33a) with *be going to* suggests that an explosion is inevitable, whether or not anyone goes near the parcel, whereas (33b) with *will* suggests that an explosion is contingent on somebody approaching the parcel:

(33)   a.   Don't go near that parcel! It's going to explode!
   b.   Don't go near that parcel! It will explode!

Again, the overtone of inevitability with *be going to* is not cancellable and so is at least partly encoded, as in (34). The overtone of contingency with *will*, on the other hand, is inferred, since it is easily cancellable, as (35) shows:

(34) ?Don't go near that parcel! It's going to explode! But if you keep well clear it won't.
(35) It will explode whether anyone goes near it or not.

Since they are not cancellable, interpretations associated with the use of *be going to*, such as the overtones of prior intention and inevitability, must be enrichments of the semantic content of the *be going to* construction. The procedural characterization of *be going to* as a relative future marker cannot give rise, on its own, to overtones of prior intention or inevitability (we cannot infer simply from the fact that an event is yet to occur that it is inevitable), so we must look elsewhere for the semantic source of these interpretations. The model of grammaticalization proposed here predicts that certain grams will exhibit semantic retention in the form of specific semantic nuances or overtones arising from their source constructions. Interpretations such as prior intention and inevitability might therefore arise as enrichments of the semantic content of the lexical source construction of *be going to*. The stages involved in the development

of *be going to* from a lexical source are repeated here:

i.   Lexical: physical progression.   The subject is in the
                                                     process of progressing
                                                     towards a physical goal.

                     (Generalization results in ...)

ii. Lexical: general progression   The subject is in the
                                                     process of processing
                                                     towards a goal.

                     ("Inference" result in ...)

iii. Grammatical: future time reference.   The subject will, at
                                                     some future time,
                                                     achieve a goal.

It is possible that the prior intention and inevitability interpretations arise as a result of inferential enrichment of the semantics of (ii), the lexical source construction expressing general progression. If the subject is capable of volition and the event described is one over which the subject has control, "being in the process of progressing towards a goal" may involve an element of planning and intention, and hence the notion of general progression may be inferentially enriched to give rise to a prior intention interpretation. If the process of progressing towards a goal is one which, once under way, cannot be reversed (such as the counting down of a timing device on a detonator), the notion of general progression may be inferentially enriched to give rise to an interpretation of inevitability.

To summarize, the *be going to* construction now encodes both procedural information concerning future time reference, and conceptual information concerning general progression. Semantic retention arises as a result of inferential enrichment of this conceptual information, resulting in interpretations such as prior intention and inevitability. These overtones only arise in certain contexts however, and a theory of semantic retention should explain why this is the case. I shall conclude this section by briefly sketching such an account with reference to the principle of relevance. The basic principles involved are the same as those used in section 3.1 to explain what happens when a lexical expression comes to be used with the same functional load as a grammatical marker.

I propose that the procedural information encoded by *be going to* is automatically recovered and used in the utterance interpretation process, since it provides a necessary processing constraint on the interpretation of an associated conceptual representation. The conceptual information also encoded will be automatically activated (or accessed) but will not necessarily be recovered and used in the interpretation process. If, on recovering

only the procedural information encoded by *be going to*, the resulting interpretation achieves optimal relevance (adequate contextual effects for minimal processing effort) on its own, the interpretation process will cease in accordance with the criterion of consistency with the principle of relevance. If, on the other hand, the resulting interpretation does not result in adequate contextual effects, or if the context is such that the choice between *be going to* and an alternative future expression, such as *will*, could result in a different proposition being conveyed (as in the minimal pairs in examples [30] and [33]), the conceptual information also encoded by *be going to* will be recovered and inferentially enriched. When this happens, *be going to* exhibits semantic retention.

6.   Summary

There is a growing body of work within the relevance theory framework which treats grammatical markers (grams) such as modality, tense, and aspect (MTA) markers, as exponents of procedural encoding. The purpose of procedural encoding is to constrain the inferential processing of conceptual representations; in the case of MTA markers, it is the processing of conceptual representations of situations and events with respect to existential, temporal, and aspectual parameters which is constrained. Grams are distinguished from other exponents of procedural encoding in that grams constrain the construction or identification of propositional forms rather than implicatures or propositional attitudes (also termed "higher-level explicatures"). Although previous accounts of procedural encoding (such as Blakemore 1987, 1988, Haegeman 1989, Klinge 1993, Wilson and Sperber 1993, Nicolle 1997a, 1997b) have adopted a synchronic perspective, these accounts should also be compatible with diachronic evidence. In this article, I have demonstrated that a procedural account of grammatical markers is compatible with research into grammaticalization (the historical development of lexical expressions into grams). In particular, I addressed three aspects of grammaticalization.

First, was the apparently gradual nature of grammaticalization. Because it is a formally gradual process, grammaticalization had been viewed as semantically gradual also. On this view the onset of grammaticalization 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, according to which we should expect that the onset of procedural encoding would be instantaneous. In section 3, I noted that this is in fact what Givón (1991) found

in his study of phonetic iconicity (through pause length) in serial-verb constructions. Further evidence that grammaticalization can be semantically instantaneous but formally gradual was provided by a consideration of the development of the Kiswahili aspectual marker *-mesha-*. This characterization of grammaticalization 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.

The second aspect to be addressed was the question of the mechanism of semantic change underlying the onset of grammaticalization. Given that the purpose of procedural encoding is to constrain the inferences drawn from utterances, we should expect to find a connection between inferential processes of utterance interpretation and grammaticalization. In section 4, I demonstrated that the conventionalization of invited inferences appears to be the only sure mechanism of semantic change underlying grammaticalization, thus bearing out this prediction.

Finally, the claim that there is in principle nothing to prevent a single expression encoding both conceptual and procedural information accounts for the phenomenon of semantic retention, whereby a gram may continue to give rise to specific interpretations derived from the semantic content of its lexical source. In section 5, I argued that, although the *be going to* construction is a grammatical marker, some interpretations conveyed by its use in certain contexts derive from the inferential enrichment of the semantic content of its lexical source expression.

It would appear, then, that semantic characterizations of grammatical markers as exponents of procedural encoding are not only compatible with research into grammaticalization, but shed light on a number of issues in this area, in particular the relation between semantic and formal features of grammaticalization and semantic retention.

Notes

*   E-mail: s.nicolle@mdx.ac.uk

1.   Principles of least effort are invoked in other cognitively influenced approaches, such as the minimalist/program in syntax (Chomsky 1993).
2.   This assumption is implicit in Blakemore's discussion of *therefore* (Blakemore 1992: 153–154). Furthermore, a procedural account of conditionals has been proposed (Foster-Cohen and Konrad 1997: 145) in which *if* may impose processing constraints on the proposition expressed, on higher-level explicatures (that is, propositional attitudes), and on implicatures.
3.   I am grateful to an anonymous referee for bringing this observation to my attention.

4.   For a typology of conceptual and procedural encoding in relation to truth-conditional versus non-truth-conditional meaning and explicit versus implicit communication, see Wilson and Sperber (1993) and Nicolle (1997b).
5.   I am grateful to an anonymous referee for this observation.
6.   This is not a question which Givón raises, as he is primarily concerned with the role of the speaker in grammaticalization. Relevance theory, on the other hand, deals principally with utterance interpretation. According to Hopper and Traugott (1993: 65) a concern with the role of the hearer in grammaticalization correlates in the literature with a focus on morphosyntactic changes, whilst a focus on the speaker's use of language correlates with an emphasis on issues such as the lexical origins of grammaticalization, pragmatic enrichment of lexical items, and semantic changes. This article is atypical therefore in focusing on the latter phenomena whilst emphasizing the role of the hearer.
7.   This is an example of a "linguistic cycle" (Hodge 1970) in which grammatical forms with similar functions keep developing from similar sources. See Ono (1992) for a discussion of linguistic cycles in the development of perfect markers in Japanese.
8.   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-*.
9.   With regard to the latter factor, see for example Dancher and Kytö (1994) on French influence on the development of English and Southern Dutch *be going to* future constructions.
10. The conventionalization of implicature is the conventionalization of what Grice termed "conversational implicatures", and should not be confused with the Gricean notion of "conventional implicatures", which are linguistically encoded but non-truth-conditional components of meaning.
11. Hopper and Traugott (1993: 84) come to a similar conclusion when they suggest that "semanticization of conversational inferences played a major role in the development of *be going to*".
12. To say that conceptual information becomes increasingly inaccessible does not entail that grammaticalization is gradual after all, since grammaticalization is equated with the onset of procedural encoding constraining the construction of propositions, not with the loss of conceptual meaning.
13. The unacceptability of (28b) indicates that *to* in the grammaticalized use of *be going to* is syntactically distinct from *to* in the non-grammaticalized use. It is therefore correct to say that the *be going to* construction takes bare nonfinite complements.
14. The importance of *to* in the development of future meaning is discussed in Hopper and Traugott (1993: 82–83). The relation between allatives and future time reference can be illustrated by comparing allative *to* with the gerundive *-ing*:

(i)   You must remember to lock the door.
(ii)   You must remember locking the door.

In (i) the event [you lock the door] is envisaged subsequent to the time of utterance, whilst in (ii) this event is envisaged prior to the time of utterance. I am grateful to David Adger for bringing this distinction to my attention.

15. I have presented formal and semantic evidence that *be going to* should be treated as a grammaticalized future tense marker; against this claim, Langacker (1990: 23–24) suggests that the flexibility of *be going to* with regard to temporal reference points together with the fact that the construction inflects indicates that it is not yet a tense marker. If Langacker is correct, the English perfect marker, *have*+past participle, is

not a grammatical marker either, as it exhibits a similar degree of flexibility with respect to temporal and conditional reference (*She will/may/would/is going to/*etc. *have eaten*) and inflects (*She has/had eaten*).

16. The following examples are taken from Nicolle (1997a) in which *be going to* is characterized as purely conceptual and therefore as a lexical expression rather than a gram; in the light of the above considerations, this characterization is incorrect.

References

Ariel, Mira
    1988   Referring and accessibility. *Journal of Linguistics* 24, 65–87.
Ashton, E. O.
    1947   *Swahili Grammar*. 2nd edition. London: Longman.
Blakemore, Diane
    1987   *Semantic Constraints on Relevance*. Oxford: Basil Blackwell.
    1988   "So" as a constraint on relevance. In Kempson, Ruth M. (ed.), *Mental Representations: The Interface Between Language and Reality*. Cambridge: Cambridge University Press, 183–195.
    1992   *Understanding Utterances*. Oxford: Basil Blackwell.
Bradshaw, J.
    1982   *Word Order Change in Papua-New Guinea Austronesian Languages*. Unpublished Ph.D. dissertation, University of Hawaii.
Bybee, Joan L.
    1988   Semantic substance vs. contrast in the development of grammatical meaning. In *Proceedings of the 14th Annual Meeting of the Berkeley Linguistics Society: Parasession on Grammaticalization*. Berkeley, CA: Berkeley Linguistics Society, 247–264.
Bybee, Joan L. and W. Pagliuca
    1987   The evolution of future meaning. In Ramat, A. G., O. Carruba, and G. Bernini (eds.), *Papers from the 7th International Conference on Historical Linguistics*. Amsterdam: John Benjamins, 109–122.
Bybee, Joan L., W. Pagliuca, and R. D. Perkins
    1991   Back to the future. In Traugott, Elizabeth C. and Bernd Heine (eds.), *Approaches to Grammaticalization. Vol.2*. Amsterdam: John Benjamins, 17–58.
Bybee, Joan L., R. D. Perkins, and W. Pagliuca
    1994   *The Evolution of Grammar: Tense, Aspect and Modality in the Languages of the World*. Chicago: University of Chicago Press.
Carey, Kathleen
    1990   The role of conversational implicature in the early grammaticalization of the English Perfect. In *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society: Parasession on the Legacy of Grice*. Berkeley, CA: Berkeley Linguistics Society, 371–380.
Carston, Robyn
    1988   Implicature, explicature and truth-theoretic semantics. In Kempson, Ruth M. (ed.), *Mental Representations: The Interface Between Language and Reality*. Cambridge; Cambridge University Press, 155–181.
    1993   Conjunction, explanation and relevance. *Lingua* 90, 27–48.
Chomsky, Noam
    1993   A minimalist program for linguistic theory. In Hale, Kenneth and S. J. Keyser (eds.) *The View from Building 20*. Cambridge, MA: MIT Press, 1–52.

Coates, Jennifer
    1983   *The Semantics of the Modal Auxiliaries*. London: Croom Helm.
Crowley, Terry
    1987   Serial verbs in Paamese. *Studies in Language* 11, 35–84.
Dancher, Andrei and Merja Kytö
    1994   The construction *be going to*+infinitive in Early Modern English. In Kastovsky, Dieter (ed.), *Studies in Early Modern English*. Berlin: Mouton de Gruyter, 59–77.
Fodor, Jerry A.
    1983   *The Modularity of Mind*. Cambridge, MA: MIT Press.
Foster-Cohen, Susan and Erika Konrad
    1997   "If you'd like to burn your mouth feel free": A relevance theoretic account of conditionals used to children. In Groefsema, Marjolein (ed.), *Proceedings of the University of Hertfordshire Relevance Theory Workshop 1995*. Chelmsford: Peter Thomas and Associates, 140–148.
Fries, Charles C.
    1927   The expression of the future. *Language* 3, 87–95.
Givón, Talmy
    1975   Serial verbs and syntactic change: Niger-Congo. In Li, Charles N. *Word Order and Word Order Change*. Austin: University of Texas Press, 47–112.
    1991   Serial verbs and the mental reality of "event": grammatical vs. cognitive packaging. In Traugott, Elizabeth C. and Bernd Heine (eds.), *Approaches to Grammaticalization, Vol. 1*. Amsterdam: John Benjamins, 81–127.
Groefsema, Marjolein
    1992   *Processing for Relevance*. Unpublished Ph.D. thesis, University College London.
Haegeman, Liliane
    1989   "Be going to" and "will": A pragmatic account. *Journal of Linguistics* 25, 291–317.
Hauner, Magdalena
    1984   Complex tenses and style in the novels of E. Kezilahabi. In Maw, Joan and D. Parkins (eds.), *Swahili Language and Society*. Vienna: University of Vienna, 107–125.
Heine, Bernd, U. Claudi, and F. Hünnemeyer
    1991   From cognition to grammar—evidence from African languages. In Traugott, Elizabeth C. and Bernd Heine (eds.), *Approaches to Grammaticalization, Vol. 1*. Amsterdam: John Benjamins, 149–187.
Hodge, Carleton
    1970   The linguistic cycle. *Language Sciences* 13, 1–7.
Hopper, Paul J.
    1991   On some principles of grammaticization. In Traugott, Elizabeth C. and Bernd Heine (eds.), *Approaches to Grammaticalization, Vol.1*. Amsterdam: John Benjamins, 17–35.
Hopper, Paul J. and Elizabeth C. Traugott
    1993   *Grammaticalization*. Cambridge: Cambridge University Press.
Jackendoff, Ray S.
    1983   *Semantics and Cognition*. Cambridge, MA: MIT Press.
Johnson-Laird, Philip N.
    1983   *Mental Models*. Cambridge: Cambridge University Press.
Klinge, Alex
    1993   The English modal auxiliaries: from lexical semantics to utterance interpretation. *Journal of Linguistics* 29, 315–357.

Langacker, Ronald W.
    1990   Subjectification. *Cognitive Linguistics* 1, 5–38.
Lehmann, Christian
    1995   *Thoughts on Grammaticalization*. Munich: LINCOM EUROPA.
    [1982]   [Originally published as *Thoughts on Grammaticalization: A Programmatic Sketch*,
                Vol. 1. University of Cologne: Arbeiten de Kölner Universalienprojekts 49.]
Li, Charles N. and Sandra A. Thompson
    1979   Third person pronouns and zero-anaphora in Chinese discourse. In Givón,
                Talmy (ed.), *Syntax and Semantics 12: Discourse and Syntax*. New York:
                Academic Press, 311–335.
Lichtenberk, Frantisek
    1991   On the gradualness of grammaticalization. In Traugott, Elizabeth C. and
                Bernd Heine (eds.), *Approaches to Grammaticalization, Vol. 1*. Amsterdam:
                John Benjamins, 37–80.
Lightfoot, David
    1991   *How to Set Parameters: Arguments from Language Change*. Cambridge, MA:
                MIT Press.
Maw, Joan and John Kelly
    1975   *Intonation in Swahili*. London: School of Oriental and African Studies.
McMahon, April M. S.
    1994   *Understanding Language Change*. Cambridge: Cambridge University Press.
Meillet, Antoine
    1912   L'évolution des formes grammaticales. In Meillet, Antoine *Linguistique
                Historique et Linguistique Générale*. Paris: Champion, 131–148.
Nicolle, Steve
    1997a   A relevance theoretic account of *be going to*. *Journal of Linguistics* 33,
                355–377.
    1997b   Conceptual and procedural encoding: criteria for the identification of linguisti-
                cally encoded procedural information. In Groefsema, Marjolein (ed.), *Proceed-
                ings of the University of Hertfordshire Relevance Theory Workshop 1995*.
                Chelmsford: Peter Thomas and Associates, 47–56.
Nurse, Derek
    1989   Change in tense and aspect: Evidence from Northeast Coastal Bantu lan-
                guages. In Haïk, I. and L. Tuller (eds.), *Current Approaches to African
                Linguistics, Vol. 6*. Dordrecht: Foris, 277–297.
Ono, Tsuyoshi
    1992   The grammaticization of the Japanese verbs *oku* and *shimau*. *Cognitive
                Linguistics* 3, 367–390.
Partee, Barbara
    1973   Some structural analogies between tenses and pronouns in English. *Journal
                of Philosophy* 70, 601–609.
    1984   Nominal and temporal anaphora. *Linguistics and Philosophy* 7, 243–286.
Pawley, Andrew
    1987   Encoding events in Kalam and English: Different logics for reporting
                experience. In Tomlin, R. S. (ed.), *Coherence and Grounding in Discourse*.
                Amsterdam: John Benjamins, 329–360.
Polomé, Edgar C.
    1967   *Swahili Language Handbook*. Washington DC: Centre for Applied Linguistics.
Recanati, Francois
    1989   The pragmatics of what is said. *Mind and Language* 4, 295–329.

Roberts, Craige
    1989   Modal subordination and pronominal anaphora in discourse. *Linguistics and
                Philosophy* 12, 683–721.
Sanders, José and Wilbert Spooren
    1996   Subjectivity and certainty in epistemic modality: A study of Dutch epistemic
                modifiers. *Cognitive Linguistics* 7, 241–264.
Schwenter, Scott A.
    1994   The grammaticalization of an anterior in progress: Evidence from a
                penninsular Spanish dialect. *Studies in Language* 18, 71–111.
Sperber, Dan, Francesco Cara, and Vittorio Girotto
    1995   Relevance theory explains the selection task. *Cognition* 57, 31–95.
Sperber, Dan and Deirdre Wilson
    1986   *Relevance: Communication and Cognition*. Oxford: Basil Blackwell.
    1995   *Relevance: Communication and Cognition*. 2nd edition. Oxford: Basil Blackwell.
Swinney, David A.
    1979   Lexical access during sentence comprehension; (re)considerations of context
                effects. *Journal of Verbal Learning and Verbal Behaviour* 18, 645–659.
Traugott, Elizabeth C.
    1988   Pragmatic strengthening and grammaticalization. In *Proceedings of the
                14th Annual Meeting of the Berkeley Linguistics Society: Parasession on
                Grammaticalization*. Berkeley, CA: Berkeley Linguistics Society, 406–416.
Ultan, Russell
    1978   The nature of future tenses. In Greenberg, Joseph H. (ed.), *Universals of
                Human Language, Vol 3: Word Structure*. Stanford: Stanford University
                Press, 83–123.
Warner, Anthony
    1993   *English Auxiliaries: Structure and History*. Cambridge: Cambridge University
                Press.
Wilson, Deirdre and Dan Sperber
    1981   On Grice's theory of conversation. In Werth, Paul (ed.), *Conversation and
                Discourse*. London: Croom Helm, 155–178.
    1988   Representation and relevance. In Kempson, Ruth M. (ed.), *Mental Represen-
                tations: The Interface Between Language and Reality*. Cambridge: Cambridge
                University Press, 133–153.
    1993   Linguistic form and relevance. *Lingua* 90, 1–25.

Key to Kiswahili sources

MWK   Abdullah, Muhammed S. (1960). *Mzimu wa Watu wa Kale*. Nairobi: Kenya
        Literature Bureau.
RM   Kezilahabi, Euphrase (1988). *Rosa Mistika*. 2nd edition. Dar es Salaam: Dar es
        Salaam University Press.