working papers in functional grammar

wpfg no. 79
August 2004

Morphology in Functional Discourse Grammar

Edited by Kees Hengeveld
University of Amsterdam

Introduction

This collection of short papers is devoted to the treatment of a number of morphological issues in Functional Discourse Grammar (FDG). The first three contributions are concerned with the characterization of morphological types in terms of the interaction between the pragmatic and semantic components on the one hand and the morphosyntactic and phonological ones on the other: Kees Hengeveld gives a general overview of the treatment of morphological types in FDG, Maria Mos then discusses a number of morphemes in the agglutinating language Tarma Quechua in more detail, and Marleen van de Vate does the same for a number of grammatical particles in the isolating language Saramaccan. The last paper, by Suzanne Dikker, goes into the problem of the treatment of semantic versus grammatical agreement within the context of FDG.

The papers in this volume are the outcome of a research tutorial organized within the context of the Research Master in Linguistics of the Graduate School for Humanities of the University of Amsterdam.

Contents

1. Morphological types in Functional Discourse Grammar

Kees Hengeveld
Department of Theoretical Linguistics, University of Amsterdam

1.1. Introduction

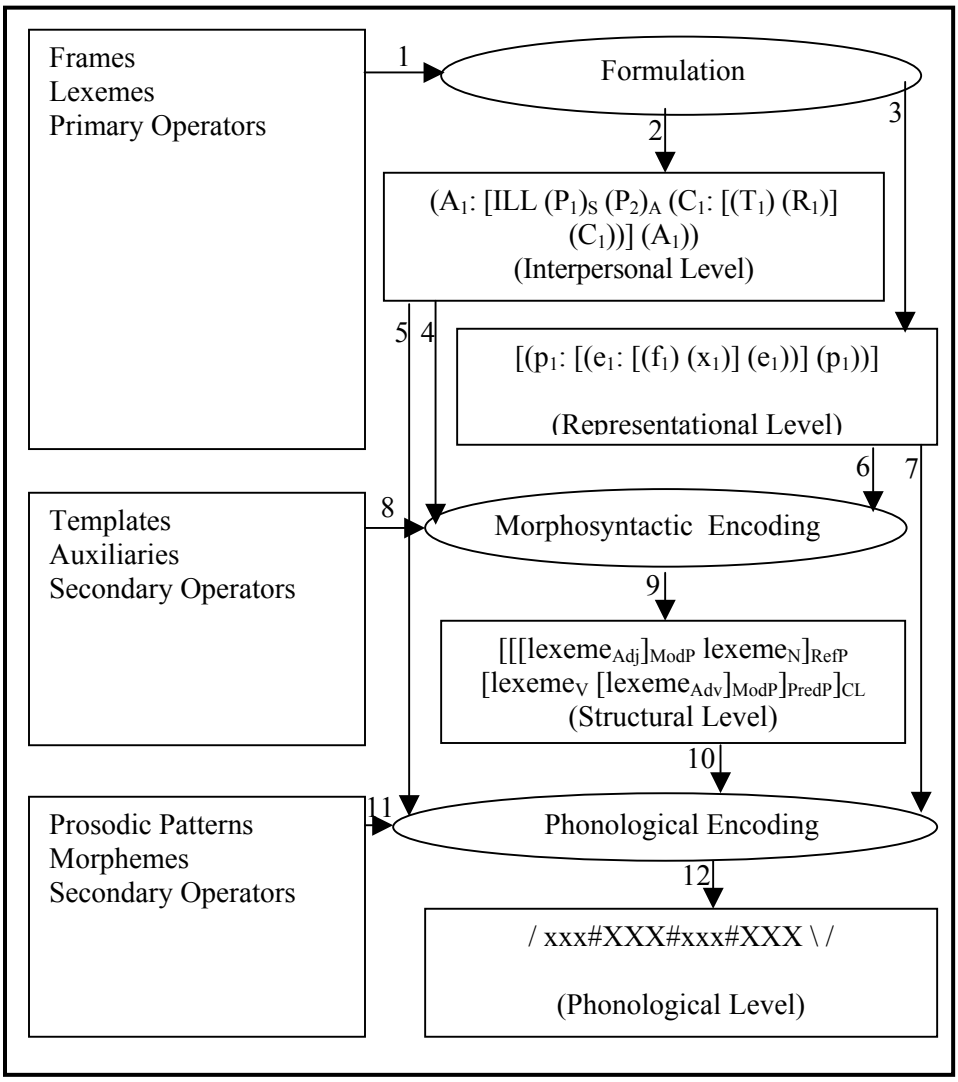
The aim of this paper is to give an overview of the way in which the differences between morphological types can be accounted for systematically in terms of the various modules that make up the model of Functional Discourse Grammar. After summing up the most relevant aspects of FDG in section 1.2, a classification of morphological types is given in section 1.3. Section 1.4. then reviews the way in which FDG handles each of these types separately. Section 1.5. summarizes the paper.

1.2. Outline of the FDG model

Figure 1 gives a general overview of the FDG model. A summary of the various properties of this model may be found in Hengeveld (forthcoming); a full presentation of the model will be given in Hengeveld & Mackenzie (in preparation). An important property of the model in the context of the present discussion is that it distinguishes an interpersonal, a representational, a structural, and a phonological level of linguistic organization, and that each of these levels is built up using different sets of primitives. The interpersonal and representational levels of organization are structured on the basis of pragmatic and semantic frames, into which lexemes and primary operators (i.e. operators that are defined in terms of their meaning) are inserted. The structural level is organized in terms of morphosyntactic templates, into which, apart from lexical material from the preceding levels, grammatical words and morphosyntactic secondary operators

(i.e. operators anticipating bound grammatical expressions) are inserted. The phonological level, finally, is organized in terms of prosodic patterns, into which, apart from the lexical and grammatical words from the preceding levels, bound morphemes and phonological secondary operators (i.e. those anticipating acoustic effects of certain morphosyntactic configurations) are inserted.

Figure 1. Outline of FDG



Free grammatical morphemes have to be introduced at the structural level, since, unlike bound grammatical morphemes, they occupy slots in the syntactic configuration, which is determined at this level. Bound grammatical morphemes are introduced at the phonological level since in many languages the form of grammatical morphemes may be affected by the syntactic configuration in which they occur. Therefore morphosyntactic secondary operators are inserted at the structural level in the appropriate position, anticipating the morphological means of expression that will eventually be selected at the phonological level.

1.3. Morphological types

Morphological types can be defined along two parameters: semantic transparency, and synthesis. Along the first parameter one can distinguish isolating, agglutinating, and fusional languages. Isolating languages are semantically transparent in the sense that there is a one-to-one relation between a word and a unit of meaning, whereas in agglutinating languages there is a one-to-one relation between a morpheme and a unit of meaning. Fusional languages are semantically opaque, in the sense that there is no one-to-one relation between a unit of form and a unit of meaning. Along the second parameter one may distinguish between polysynthetic and non-polysynthetic languages. Polysynthetic languages allow the presence of more than one lexical element within a single word, non-polysynthetic languages do not. The two parameters are basically independent of one another: the first has to do primarily with the status of grammatical elements in the language, the second one with the status of lexical elements. As a result, polysynthetic languages can be fusional or agglutinating just like non-polysynthetic languages. The only restriction in terms of combinations of the two parameters is that a polysynthetic language cannot at the same time be isolating. Note furthermore that many languages exhibit features of more than one morphological type.

Examples from languages from these different types are given below. Fijian (1) is an isolating language, Turkish (2) an agglutinating language, Spanish (3) a fusional

language, and Southern Tiwa (4) a polysynthetic language. The glosses clearly reveal the morphological structure of the languages involved: in (1) the gloss is a word-by-word translation, in (2) a morpheme-by-morpheme translation, in (3) a one-to-many translation, and in (4) the gloss reveals the presence of two lexical elements. Note that (4) is a case of syntactic incorporation, as the incorporated object is cross-referenced on the verb.

Fijian (Milner 1972: 42)

(1) Mo dou kauta mada yani na cina.
IMP 2PAUC take MIT away ART lamp
'Take the lamp away.'

Turkish (van Schaaik p.c.)

(2) Anlı-y-abil-ecek-miş-im.
understand-y-ABIL-IRR-INFER-1.SG
'I gather I will be able to understand.'

Spanish

(3) Lleg-ó.
arrive-IND.PAST.PF.3.SG
‘He/she/it arrived.’

Southern Tiwa (Gerdts 1998: 88)

(4) Te-shut-pe-ban
1.SG>PL-shirt-make-PAST
'I made (the) shirts.'

1.4. The representation of morphological types in FDG

The application of the FDG model, and particularly its division of labour between the various components, to the examples just given, leads to the following analyses of the examples just given.

Figure 2 contains the analysis of the Fijian example (1). It shows that the grammatical words *mo*, *mada*, and *na* are inserted at the structural level, and furthermore can be seen as the direct translation of the basic illocution 'IMP' and the primary operators 'MIT' and 'SPEC', thus reflecting the semantic transparency of this type of language at the syntactic level. For a detailed analysis of the isolating language Saramaccan see chapter 3.

Figure 2. Fijian

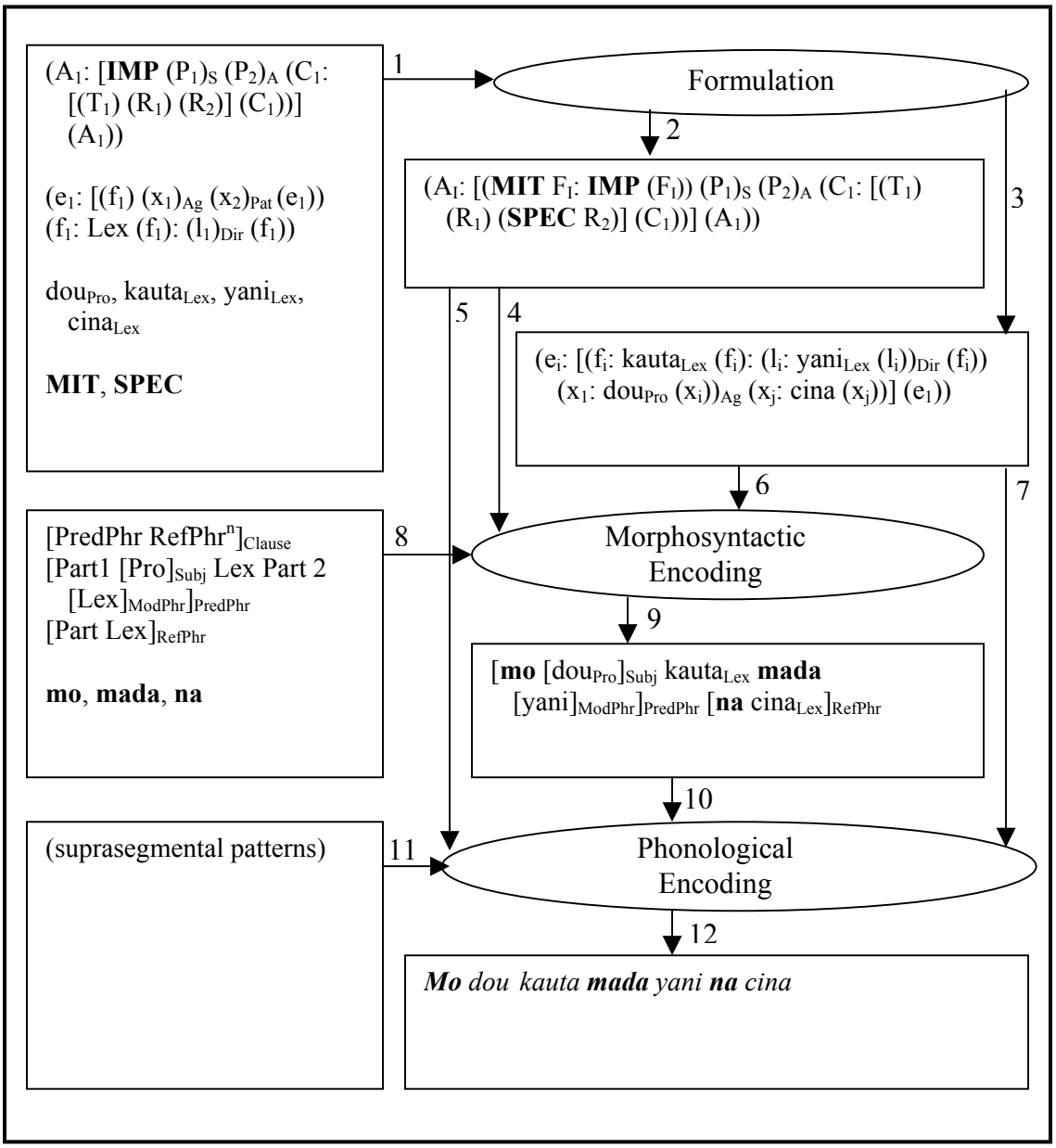


Figure 3 contains the analysis of the Turkish example (2). It shows that the bound grammatical morphemes *-AbIl*, *-EcEk*, *-mIş*, and *-Im* are inserted at the phonological level, as the expression of the secondary operators 'Abil', 'Irr', 'Infer', '1Sg' that are introduced at the structural level. Note that the secondary operators are in a one-to-one relationship with the primary operators 'abil', 'irr', 'infer' and the person marker '1sg', thus reflecting the semantic transparency of this language at the morphological level. For a detailed analysis of the agglutinating language Tarma Quechua see chapter 2.

Figure 1.3. Turkish

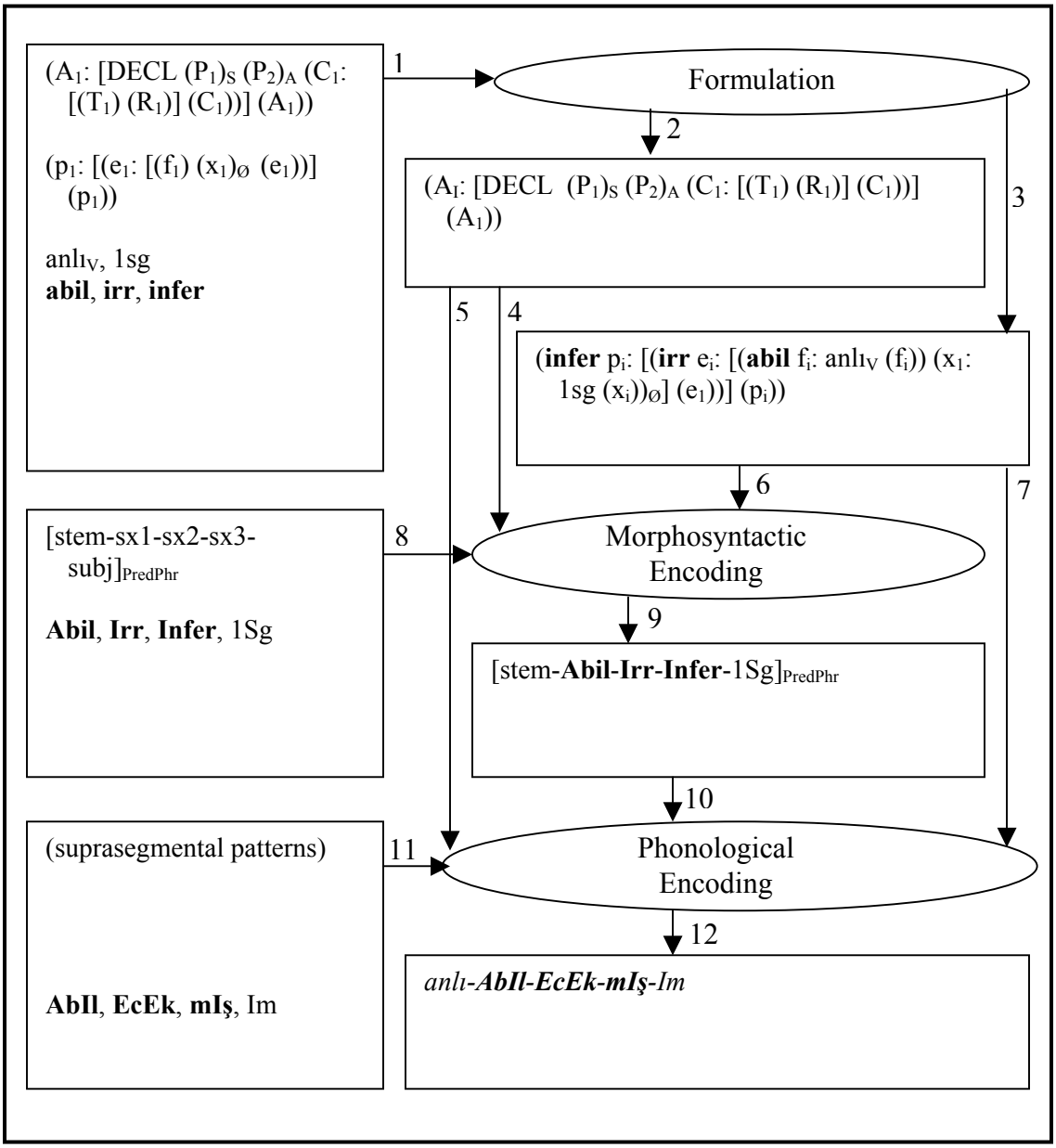


Figure 4 contains the analysis of the Spanish example (3). It shows that the bound grammatical morpheme *-ó* is inserted at the phonological level, as the expression of the secondary operator 'IndPastPf3Sg' that is introduced at the structural level. In contrast to the previous case, the selection of this secondary operator is triggered by the joint presence of the basic illocution 'DECL', the primary operators 'past' and 'pf', and the person marker '3sg', thus reflecting the lack of semantic transparency of this language within its inflectional system.

Figure 4. Spanish

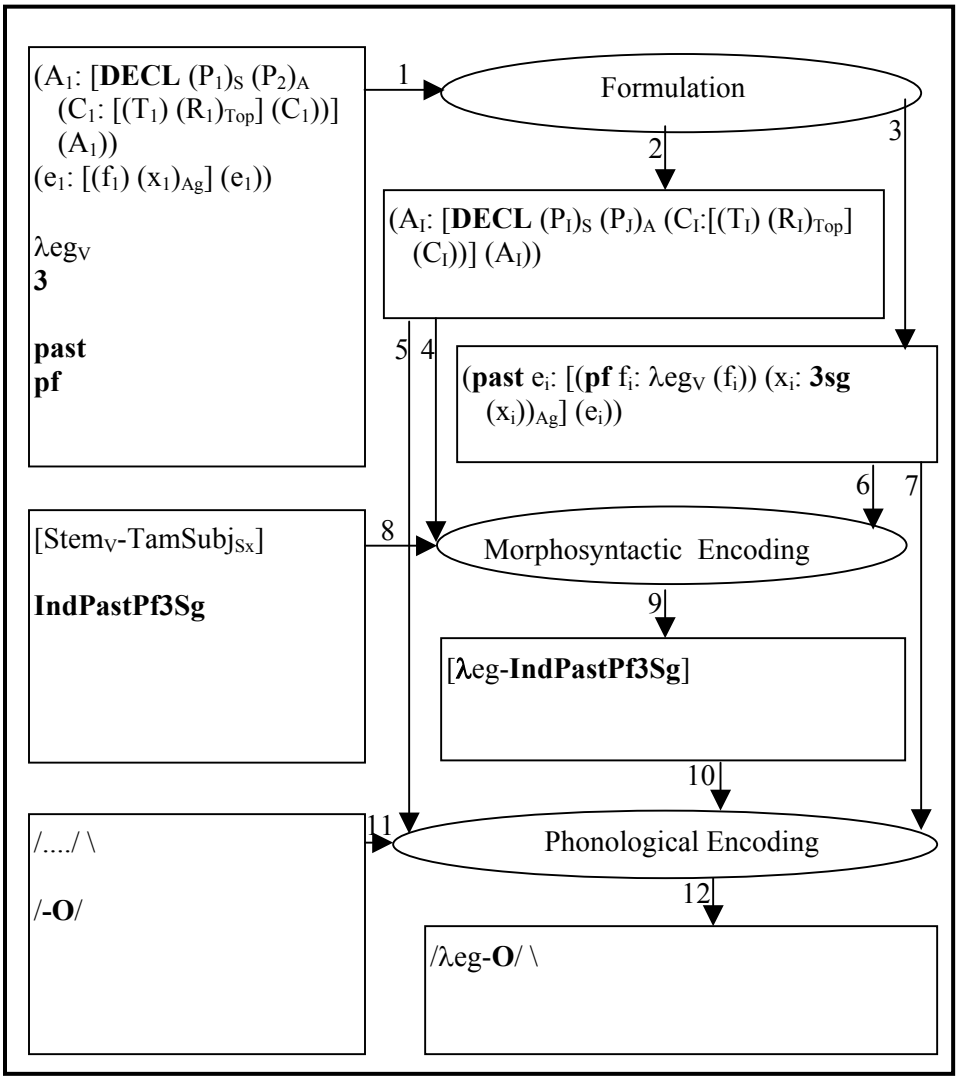
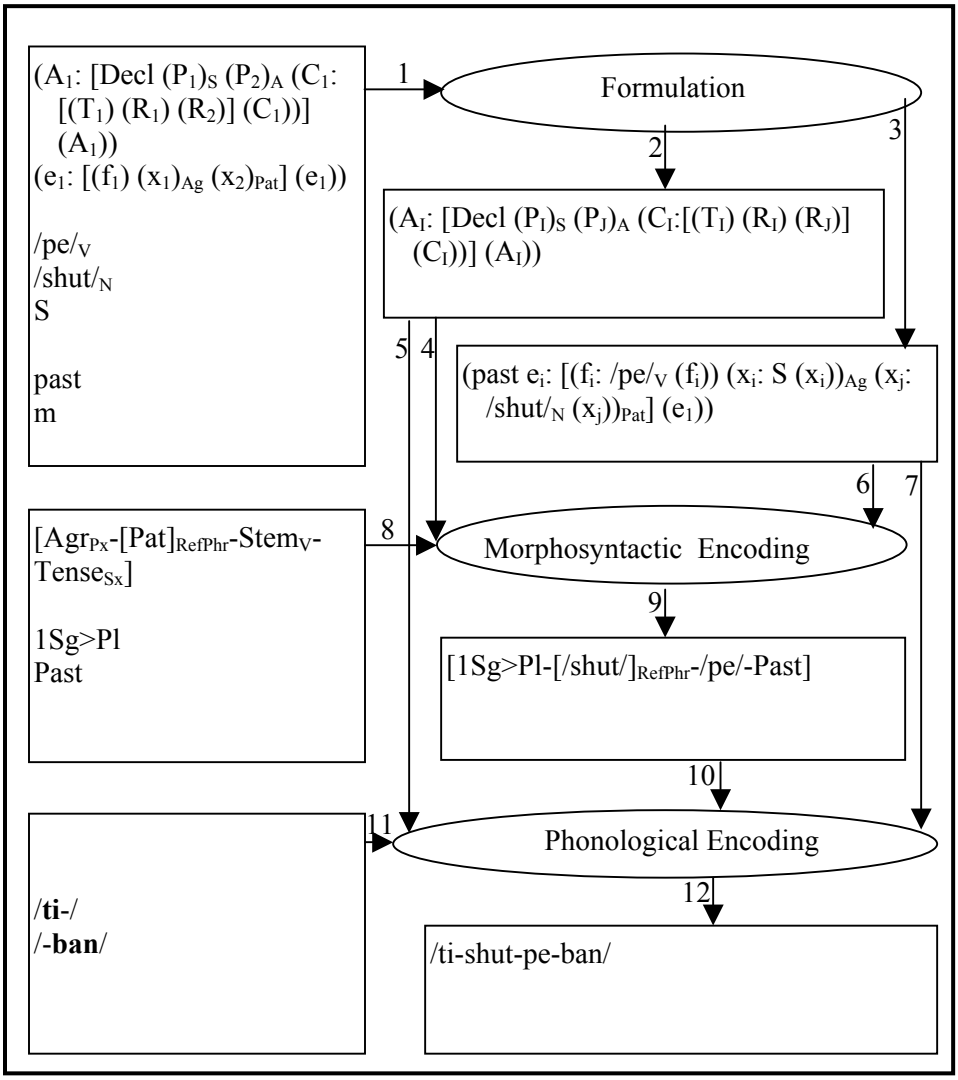


Figure 5 contains the analysis of the Southern Tiwa example (4). Apart from the agglutinating characteristics of this example, the analysis shows that two lexically filled units at the representational level, (f_i) and (x_j), are inserted into a single complex word template at the structural level, thus reflecting the semantic independence of the two units on the one hand, and their structural dependency on the other.

Figure 5. Southern Tiwa



1.5. Summary

Figure 6 summarizes the relationship between the various levels in FDG in languages of distinct morphological types along the parameter of semantic transparency. Recall that in this paper this parameter was used to characterize the differences between languages as regards grammatical forms, i.e. grammatical words and bound morphemes. These elements show no isomorphism across levels of representation in fusional languages, isomorphism between the interpersonal/representational levels and the structural level in isolating languages (in which the phonological level is irrelevant for the expression of grammatical elements), and isomorphism between all levels in agglutinating languages.

Figure 6. Semantic transparency

<table>
<tr><th>Morphological type</th><th>Interpersonal/ representational level</th><th>Structural level</th><th>Phonological level</th></tr>
<tr><td>Fusional</td><td></td><td></td><td></td></tr>
<tr><td>Isolating</td><td colspan="2">isomorphism</td><td></td></tr>
<tr><td>Agglutinating</td><td colspan="3">isomorphism</td></tr>
</table>

Figure 7 summarizes the relationship between the various levels in FDG in languages of distinct morphological types along the parameter of semantic transparency. Recall that in this paper this parameter was used to characterize the differences between languages as regards the use of lexical forms. In polysynthetic languages there is isomorphism across the structural and phonological level of analysis, but there is no one-to-one relation between units at these levels and units at the interpersonal and/or representational ones. In non-polysynthetic languages there is such a one-to-one relation across the various levels.

Figure 7. Synthesis

Morphological type	Interpersonal/ representational level	Structural level	Phonological level
Polysynthetic		isomorphism	
Non-polysynthetic	isomorphism		

The FDG model thus helps to arrive at a systematic characterization of languages as regards their morphological types in terms of differences in their underlying representations.

References

Gerdts, Donna B. (1998), Incorporation. In: Andrew Spencer and Arnold M. Zwicky eds, *The Handbook of Morphology*, 84-100. Oxford: Blackwell.

Hengeveld (forthcoming), Dynamic expression in Functional Discourse Grammar. In: Casper de Groot & Kees Hengeveld eds, *Morphosyntactic expression in Functional Grammar* (Functional Grammar Series 27). Berlin: Mouton de Gruyter.

Hengeveld, Kees & J. Lachlan Mackenzie (in preparation), *Functional Discourse Grammar*. Oxford: Oxford University Press.

Milner, G.B. (1972), *Fijian grammar*. Suva, Fiji: Government Press.

2. Tarma Quechua morphology in the FDG-model

Maria Mos
Graduate School for Humanities, University of Amsterdam

2.1. Introduction

Tarma Quechua has various affixes that may either have scope over a whole clause or mark contrasts between different constituents within the clause. These affixes can be attached indiscriminately to both nominal and verbal constituents. The challenge this feature of Quechua morphology poses to the FDG model is clear: how can these affixes be represented in the underlying representation? Which affixes are triggered at which levels? How can we do justice to the elasticity of these affixes? If an affix can be added to nominal or verbal constituents, is it possible to generalize across these uses in underlying morphosyntactic templates?

In this paper, I will argue that the FDG model can indeed account for the occurrence of these markers. In the first paragraph the relevant language data are introduced and illustrated on the basis of the first three lines of a story. The analysis focuses on the occurrence of two affixes: the hearsay marker -*š(i)* and the relevance marker -*q(a).* The second paragraph proposes an underlying representation for the three lines of text. In the third paragraph I go into the encoding process which produces the appropriate occurrences of -*š(i)* and -*q(a)*. The concluding paragraph consists of a short summary of the present findings and points at some remaining questions.

2.2. The data

2.2.1 Tarma Quechua

Quechua languages are morphologically very rich and strongly agglutinative. This is also true of Tarma Quechua (Adelaar 1977). Their morphological make-up makes Quechua languages very interesting for linguists interested in morphology: the structure of complex words, phrases and clauses is relatively transparent.

In the terminology of Hengeveld, Rijkhoff & Siewierska (2004), the parts-of-speech system of Quechua languages is of the flexible type 2/3. This means that in the basic lexicon there is a class of verbs and a class of non-verbs, and no specialized classes of basic adjectives and manner adverbs. When derived lexemes are taken into consideration, a specialized class of adjectives may be distinguished. Furthermore, many words can be used both as nouns and as verbs. For instance, the word *kača* can mean 'a fire' in referential use, but also 'to light a fire' in predicative use. In these cases the nature of the syntactic slot that a lexeme occupies in a particular sentence becomes clear through its position and through the morphological markers added.

2.2.2. Class Free morphological markers (constituent or clausal scope)

As stated above, there are affixes that can be added to verbal and nominal constituents alike. In Adelaar's (1977) terminology these markers are 'class free' (CF) affixes.[1] He distinguishes eight different CF affixes, each of which can take on various forms depending on the phonological environment. In the FDG model, these different forms are dealt with through morphophonological rules within the articulator, which is part of the output component.

CF suffixes are word-final; they follow all other suffixes. They are also constituent-final: when a constituent is complex, the suffix is always added to the last

[1] Lastra (1968) distinguishes a group of suffixes in Cochabamba Quechua with much the same forms and meanings and calls them 'independent suffixes'.

word of that constituent, as in (1) (Adelaar 1977: 77), where the interrogative suffix *-ču* is added twice to the last element of the two noun phrases, which in this case happens to be the head noun.

(1) *yurax uĉbataču o xarwaš uĉbataču*[2]
white powder-INT or yellow powder-INT
'[Do you want] white powder or yellow powder?'

Some CF suffixes can co-occur attached to the same word, but these possibilities are limited. In general, no more than one CF affix will be attached to a single constituent.

The scope of this paper does not allow for an exhaustive treatment of all eight CF suffixes. Instead, I will focus on the occurrence of two of them, the relevance marker *-qa* and the hearsay marker *-ši*, in the first three lines of a story in the San Pedro de Cajas dialect of Tarma Quechua (Adelaar 1977: 412):

(2) a *manseba:du ulqu-qa aywa-ra-ø biya:hi-ta-š*
adulterous man-RELV go-PST-3 trip-ACC-HEAR
'It is said that an adulterous man went on a trip'.

b *biya:hi-ta aywa-ru-pti-n-ši*
trip-ACC go-ANT-DS-3-HEAR
'As he went on that trip, ...'

c *warmi-q' ke:da-ra:-ri-n wawa-la-n-kuna-wan*
wife-RELV stay-PST-PL-3 child-RESTR-3-PL-INSTR
'... the wife stayed behind alone with her children.'

[2] In these examples words are not structurally analyzed and glossed if this is not relevant for what is intended to be shown. All examples are taken from Adelaar (1977).

2.2.3. *-qa* and *-ši*

The three lines of text that were introduced in the previous section contain two instances each of the CF broad scope affixes: *-q(a)* and *-š(i)*[3]. I will first discuss *-qa* and then turn to *-ši*.

According to Adelaar (1977:71), *-qa* is used for two clearly distinct purposes. It may be used 'to mark a segment as containing non-focal information, that is, (old) information necessary for a proper understanding of the transmitted message itself', or it may be used 'to mark a contrast between two sentences'. In the analysis of Lastra (1968: 41) *-qa* is a marker that 'indicates which word is the topic of the utterance'. Weber (1989: 404), discussing Huallaga Quechua, claims that '*-qa* occurs on those constituents of a sentence the speaker wishes to indicate as most responsible for the sentence's relevance to its context', where by 'context' he means 'the set of propositions that the speaker assumes the hearer to know at the point at which he says the sentence'. This is a rather more general definition than 'topic', the usual gloss of the affix, and one which more or less captures the occurrences of the suffix in (2). I will use the gloss RELV here to signal its relevance-marking meaning. The position of the marker is clear: word- and constituent-final. Note that the scope of the notion RELV is phrasal.

The affix *-ši* is one of the markers that belong to Weber's class of *evidential suffixes*. These suffixes indicate how the speaker has acquired the knowledge that he is transmitting in his utterance. This particular affix conveys that the speaker's source of information is hearsay (hence the gloss HEAR). According to Weber, this affix is especially common in folktales, and the lines in (2) are indeed taken from such a folktale. The HEAR marker is usually attached to the first constituent of a sentence. If that constituent is already marked with another CF affix, however, the affix is attached to the next constituent to the right in the sentence. Despite the fact that the HEAR marker normally attaches to the first constituent, its scope is clausal.

[3] As shown in paragraph 1.2, the form of the suffixes may differ. In this case, the final vowels *–a* in *–qa* and *–i* in *-ši* are deleted because the suffixes occur after a short vowel.

2.3. Underlying representations

The underlying representation of sentences (2a) and (2b-c) are given in Figures 1 and 2, respectively.

Figure 1. Analysis of example (2a)

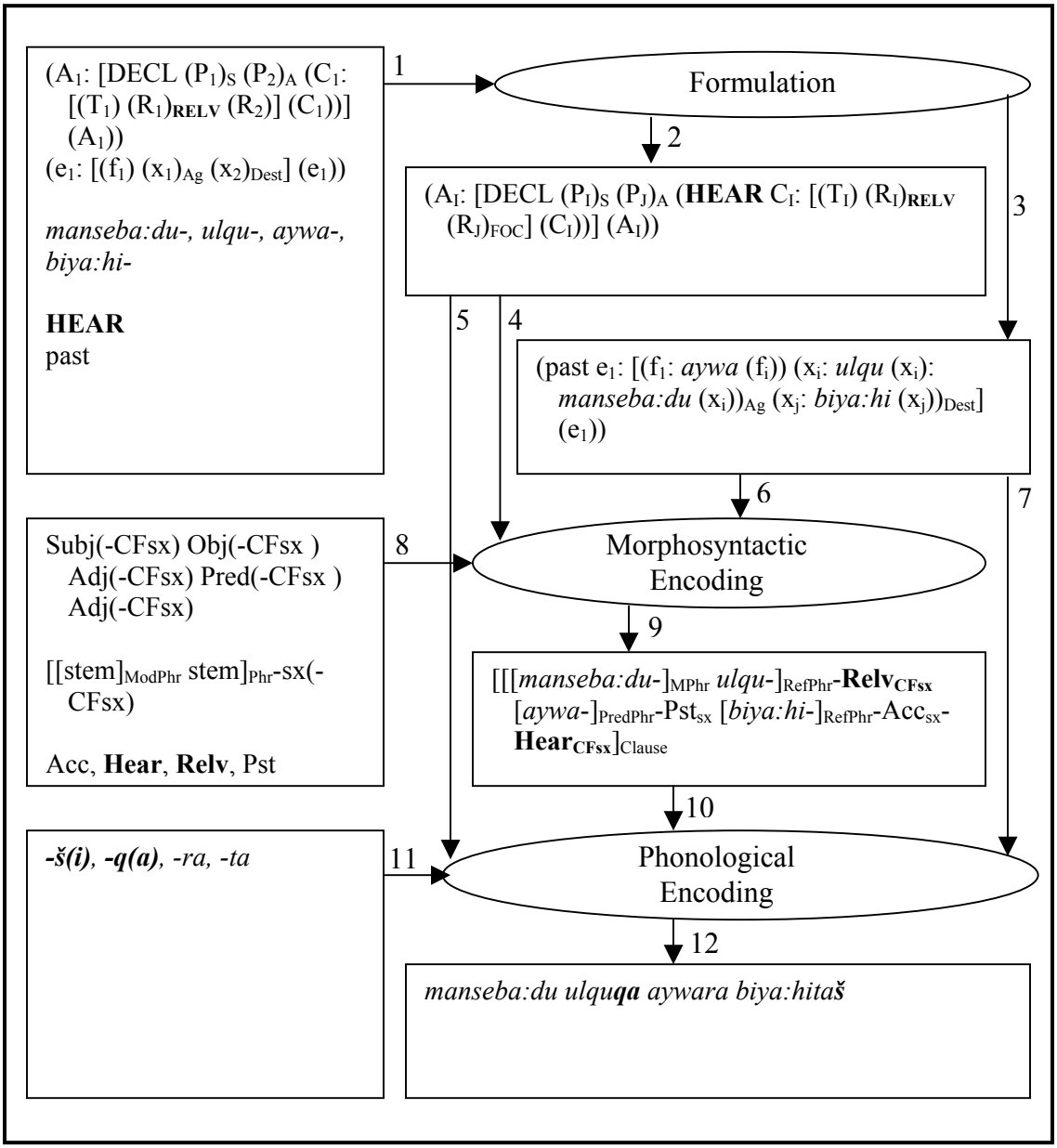
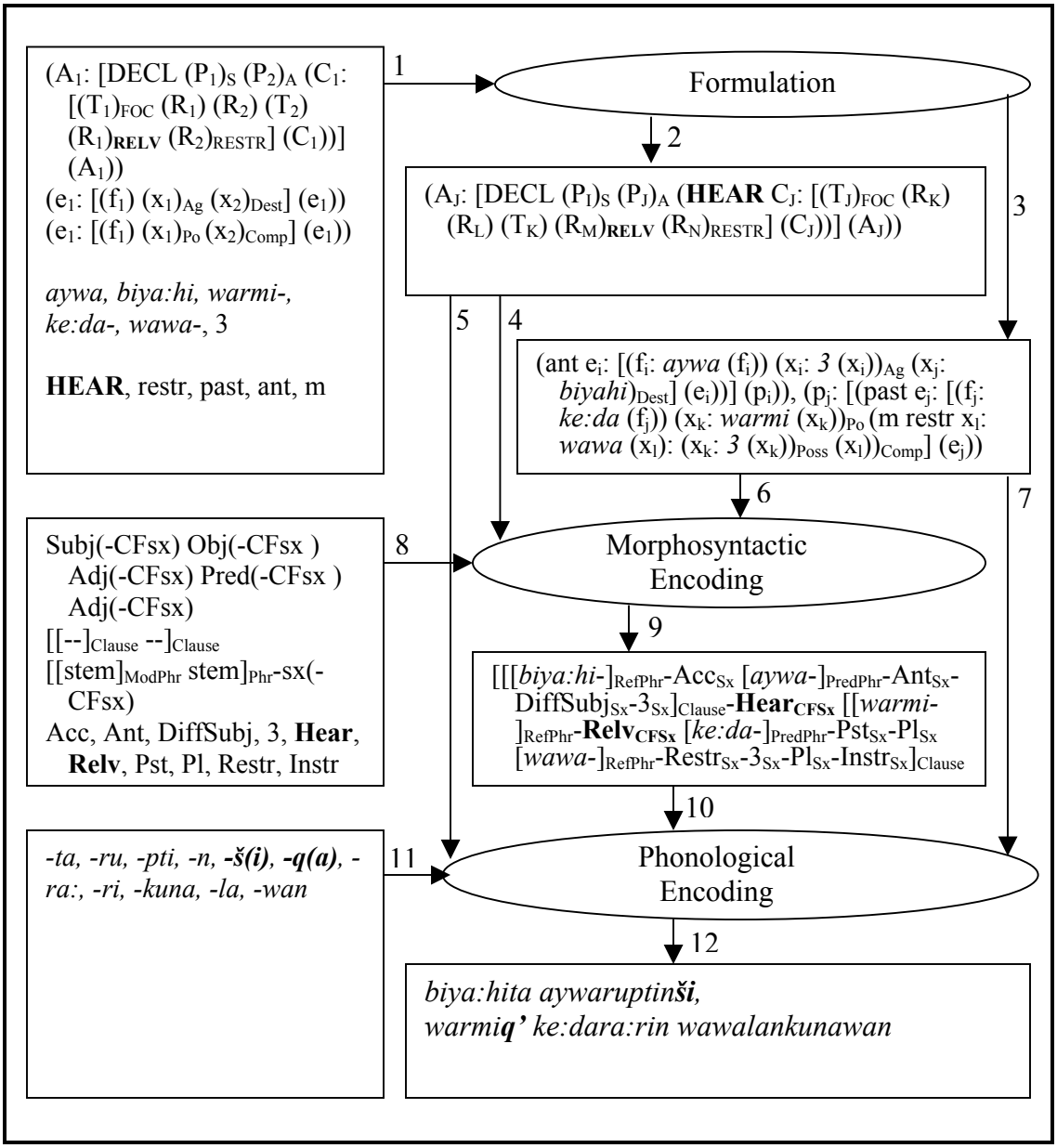


Figure 2. Analysis of example (2b-c)



The analysis is restricted in two ways. Firstly, the word templates (briefly touched upon in paragraph 2.4.2) in both the nominal and verbal domain are highly complex because of the rich morphological marking.[4] The details of these word templates fall outside the scope of this paper. All non-CF-affixation is therefore left unanalysed. Secondly, the phonological encoding takes care of the selection of the actual form of a particular suffix, on the basis of the relevant information from the fund (arrow 11). Much more is done at this stage, like sentential intonation. I do not have sufficient information on the prosodic patterns in Tarma Quechua to represent these here.

2.4. CF affixes

2.4.1. The problem

The fund provides the formulator (arrow 1) with the relevant frames, lexical items and primary operators. Among the functions that enter into interpersonal frames is the one labelled here as RELV. Among the primary operators is HEAR. These two linguistically relevant notions enter the underlying representation at the interpersonal level (arrow 2). The referential acts R_I and R_M carry the RELV functions since they introduce descriptions of entities that make the sentences in which they are described relevant for the story. HEAR denotes that the speaker has come to know this information through hearsay. There is no linear ordering in these levels yet. The ordering of the words and phrases is only determined later, deeper down in the model.

At the morphosyntactic level, the fund supplies the encoder with the relevant templates at the clause, phrase and word levels (which will be discussed in paragraph 2.4.2), and with secondary operators. The templates provide slots for phrases, words, and affixes. The interpersonal function RELV, with its phrasal scope, is morphosyntactically encoded as an affix at the end of the constituent that expresses the relevant entity. Consequently, that constituent is no longer available for the attachment of the clausal CF marker HEAR. This is the problem I want to consider here.

[4] See for example Muysken (1986) for a proposal concerning the verb template.

The clausal CF suffix HEAR is, as Adelaar (1977: 79) notices, usually attached to the first constituent of a sentence. This is illustrated in (2b-c), represented in Figure 2, in which the affix marker HEAR is attached to the embedded clause, which is the first constituent of the main clause. If the first constituent of a clause is marked with another CF affix, however, the HEAR marker cannot be expressed on that constituent any longer, since-*ši* seems to be mutually exclusive with other CF suffixes, which points to the occupation of a common slot in the phrase template. As Adelaar (1977: 80) notes, in these cases the HEAR marker prefers to be attached to a focal constituent, if any. This is arguably the case in (2a) as represented in Figure 1.

How can the FDG model reflect such a constraint on the ordering of CF suffixes, if there is no ordering at the higher levels? One of the principles in FDG, discussed by Hengeveld (forthcoming), is the *depth first principle*. According to this principle, 'information from a certain level is sent down to a lower level as soon as the necessary input information for that lower level is complete'. The semantic notion hearsay is encoded in a suffix that can occupy the CF slot of the first constituent in the clausal template. If this notion were sent down to the morphosyntactic level before the notion ‘relevant’ is available for that first constituent, it would no longer be possible to mark the relevance of the entity encoded by that constituent. If the notion of relevance reaches the morphosyntactic level first, there is no problem, since the hearsay marker can also attach to another constituent, as specified in the clausal template. Its scope is clausal, unlike the (phrasal) relevance marker, and the marker can attach to the other constituents, even if it is more commonly found attached to the first one. But in sentence (2a) the affix *-qa*, encoding the notion 'relevant', can only be attached to the clause-initial constituent. In this case the hearsay marker has to be added to a different constituent.

2.4.2. Template selection

The fact that the possibility of marking the relevance of a constituent seems to depend on how quickly this notion is sent down is not very desirable within the FDG model. In order to provide a solution for the problem posed for the data presented earlier by the

depth first principle in combination with the single CF slot on the first constituent, it is necessary to look more closely at templates and the way they are triggered.

Templates at all levels in this model are as general as possible and as specific as necessary at the same time. In general, inflectional affixes in Quechua are not suffixed to the lexical head of a phrase but to that phrase as a whole. This allows for a very general template for phrases.[5] When the combination of pragmatic and semantic information and the expression of the syntactic relations as specified in the predicate frame leads to the selection of more than one grammatical element in the form of an affix to a stem, however, a more specific template will have to be used. For the ordering of the actual affixes, more specific word templates are available.

There are various possible solutions to the problem outlined in the previous paragraph. In order to avoid the possibility that the HEAR marker is attached to the first constituent, thereby blocking the insertion of the RELV marker, it is necessary that either

(i) the RELV marker is always introduced first; or

(ii) the clausal frame is very specific, i.e. the RELV marker is part of the clausal frame; or

(iii) the clausal frame is selected in a process in which the choice is progressively narrowed down through activation.

As for (i), this would mean a severe restriction on the depth first principle, and is therefore not very desirable. Hannay (1991) proposes a model for English declarative sentences which is along the lines of (ii). He distinguishes five different modes, e.g. the ALL-NEW mode and the REACTION mode. In English, these different modes are expressed through different word orders. For Tarma Quechua, the RELV marker would be (one of) the way(s) in which the specific mode is expressed. The weak point of applying this proposal to the data at hand, I feel, is the fact that the notion relevance is not directly relevant for the marking of HEAR. In other words: if the relevance of a non-first constituent is expressed, the hearsay marker will attach to the first constituent, and RELV need not be considered to select the right template for the encoding of HEAR.

[5] In fact, as illustrated earlier, in many cases a lexeme is ambiguous between a verbal and a nominal reading, and it is only the nature of the suffixes attached to the phrase that shows how a lexeme is used.

For this reason, I think a model of selection and activation would be more adequate. Some of the information that comes down from the interpersonal and representational levels can be expressed in more than one way. I propose that in these cases more than one template is activated. Take for instance the first sentence, which consists of a verbal predicate with two arguments. The most common ordering of these three elements in Tarma Quechua is Agent-Adjunct-Predicate. In general, the verbal predicate is the last constituent. This is not the only possible ordering, however, as is evident from the examples in (2). In neither of these is the verb the last constituent. My suggestion is that initially all templates are activated in which the elements that are sent down can be expressed. Some templates are used very often. These will therefore easily reach the threshold and be selected more readily than less common ones. Different pieces of information that are coming down may activate different subsets of templates. The notion 'relevant' will very strongly activate a phrasal template in which *-qa* is attached to the last word in the phrase describing the relevant entity. The selection of this template will be immediate, because the only way to express the notion is with this one template. The notion of 'hearsay' activates a range of clausal templates. Some of these templates will contain a verbal predicate with two arguments, and these templates will be activated by the lexemes that come down as well. Let's say that these templates become much more activated than all the other possible clausal templates. Then there are still a couple of clausal templates in each of which the marker *-ši* is added to the last word of a constituent. The most frequently used of these frames is the one in which -ši is added to the first constituent. However, the notion of relevance is also sent down, and strongly activates a template in which the CF slot of the agent, typically expressed as the first constituent in Tarma Quechua, is taken by *-qa*. The sum of the activations results in the selection of the template in which the first constituent is marked for relevance (*-qa*) and another one for hearsay (*-ši*). If one of these other constituents expresses a focal referent, this will be the one on which the hearsay marker will be expressed.

Notice that this doesn't mean that *-ši* is moved. When the selection of a template is dealt with in terms of activation levels and thresholds, there is no need for movement of any kind. The template selection procedure I propose does entail that the *depth first* principle is limited in its application: for a clausal template to be selected, it is necessary

to know the argument structure of a sentence, and who or what is the agent. It is the agent who is most relevant to the story, and therefore it is semantically and pragmatically (almost) impossible for the relevance of the agent not to reach the morphosyntactic level at a very early stage of the process. Different factors, such as topicality, can further activate specific templates. Each factor, every piece of information coming down, can influence the selection of a template to a different degree.

2.5. Summary and conclusion

The FDG model can adequately describe the Class Free affixes that occur in Tarma Quechua at the level of meaning (interpersonal and representational levels), and the position they occur in (phrase finally, and often mutually exclusive). The fact that these markers may occur on verbal and nominal constituents can be elegantly represented because phrasal templates can be specified in a very general way. In this paper I have argued that a process of activation and selection takes care of the problem posed by the fact that different markers may occupy the same slot but cannot co-occur attached to the same constituent.

Clearly, more thought needs to go into the concept of templates: what form do these take? How abstract are they? How is one selected over others, especially when more than one template seems to be able to express everything in the underlying representation? Presumably, the strength with which something that is sent down to the morphosyntactic encoder activates templates may differ. Some things may activate a template to such a degree that it is immediately selected. The importance of a specific piece of information such as relevance, i.e. the influence it has in the selection process, will be different cross-linguistically. Viewed as such, it could be hypothesized that second language learners experience interference from their mother tongue much more strongly when these strengths are different. Further research will have to be done to investigate whether this hypothesis is correct. Secondly, I want to stress that the aim of this paper was not to show conclusively that the FDG model can handle *all* large-scope morphological marking. It would be very interesting indeed to look at other languages

that also show this type of marking, but which are structurally different, or mark semantically different concepts. What I hope to have shown, however, is that the markers that were treated in this paper can be analysed clearly, adequately and in a straightforward way within the FDG model.

References

Adelaar, Willem F.H. (1977), *Tarma Quechua. Grammar, texts, dictionary*. Lisse: Peter de Ridder.

Hannay, Mike (1991), Pragmatic function assignment and word order variation in a functional grammar of English. *Journal of Pragmatics*, 16, 131-155.

Hengeveld, Kees (forthcoming), Dynamic expression in Functional Discourse Grammar. In: Casper de Groot & Kees Hengeveld eds, *Morphosyntactic expression in Functional Grammar* (Functional Grammar Series 27). Berlin: Mouton de Gruyter.

Hengeveld, Kees, Rijkhoff, Jan & Siewierska, Anna (2004). Parts-of-speech systems and word order. To appear in: *Journal of Linguistics* 40.3.

Lastra, Yolanda (1968), *Cochabamba Quechua syntax*. The Hague: Mouton.

Muysken, Pieter (1986), Approaches to affix order. *Linguistics*, 24, 629-643.

Weber, David J. (1989). *A grammar of Huallaga (Huánuco) Quechua.* Berkeley: University of California Press.

3. Saramaccan: Grammatical particles in an isolating language[6]

Marleen van de Vate
Graduate School for Humanities, University of Amsterdam

3.1. Introduction

This paper addresses the question of how Functional Discourse Grammar (FDG) deals with grammatical particles in an isolating language. This is a particularly interesting issue in the context of this grammatical model, since grammatical particles in isolating languages are the expression of a wide range of meanings and functions at the interpersonal and representational levels, but are at the same time very similar in terms of morphosyntactic behaviour. This paper will show that these two aspect of grammatical particles can be accounted for in a transparent manner within FDG. For a general discussion of FDG and morphological typology I refer to chapter 1.

The language that will be used to study this topic is Saramaccan, a Creole language spoken in Surinam. At the end of the 17th century, slaves escaped from the plantations and fled into the interior of Surinam to the Saramaka River. The slaves formed several clans, and each developed its own dialect. Nowadays the language has around 27,000 speakers living in Surinam and in The Netherlands. The lexifier-language of Saramaccan is English, but Portuguese had a very important influence on the language as well. Around 30 percent of the lexicon is Portuguese-derived. Gbe and Kikongo are the main substrate languages. The basic word order of Saramaccan is SVO.

The following examples form the basis for my analysis. In (1) the first three lines of the folktale *Tótómbòtí* 'woodpecker' are presented. This is an oral narrative about the origin of the river and the woodpecker's red crest. Sentence (1b) will be analyzed later in

[6] I am grateful to Vinije Haabo for his help in analyzing the Saramaccan data.

this paper, since it contains two grammatical particles: *hɛ̀n* and *dí*. Sentences (1a) and (1c) provide the context that is necessary for the analysis of (1b). Sentence (2) is selected because it contains a third particle: *bì*.

(1) a

hɛ̀n	*tótómbòtí*	*wákà*	*té*	*kó*	*dóù*
hɛ̀n	totómbòtí	wáka	te	kó	dóu
then	woodpecker	walk	till	come	arrive

'Then the woodpecker walked out there'.

b

hɛ̀n	*à*	*tjɔ̀kɔ̀*	*dí*	*sítónù*	*kóokóokóo*
hɛ̀n	à	tjɔ̀kɔ̀	dí	sitónu	kookookoo
then	3SG	stab	DET	rock	IDEO

'Then he pecked at the rock. Peck! Peck! Peck!'

c

hɛ̀n	*à*	*wáká*	*gó*	*sééká*	*táámpù*
hɛ̀n	à	wáka	gó	seeká	taámpu
then	3SG	walk	go	arrange	stand

'Then he walked to [another place] and got himself ready'.

(2)

à	*bì*	*jáka*	*en*	*púu*
à	bì	jáka	en	púu
3SG.NOM	PAST	chase	3SG.OBJ	remove

'He chased him away'.

3.2. Some preliminary issues

Before turning to the analysis of grammatical particles in Saramaccan, I will briefly go into a number of other properties of Saramaccan that show up in the examples, since these will have to be taken into account within the FDG representations that will be presented in section 4.

Saramaccan is a tone language. At the tonemic level, three tones can be distinguished: high tones, low tones, and unspecified tones. The unspecified tones adapt to their environment: if they are surrounded by high tones they will surface as high tones, in all other situations they will surface as low tones. This adaptation of the tones to their phonological environment is taken care of within the output component of the overall FDG model, i.e. outside the grammar proper. In example (1)-(2) the first line contains the tones in their phonetic realization, and the second line the underlying representation of the tones.

A second aspect that merits attention is the absence of tense markers in (1). The three lines of text in this example are taken from an oral narrative. The transcription of the narrative is a direct reflection of the way the narrator told the folktale to his audience. This has certain consequences for the interpretation of these lines. In the first line of the story (not examined here),[7] the narrator situates the first event in the past. From this moment onwards no further tense markers have to be used in the narrative. The meaning of a sentence without a tense marker depends, in Saramaccan, on the kind of verb. Stative verbs without a tense marker have a present tense reading, non-stative verbs a simple past reading. When *bì* is present in a sentence with a stative verb, this will lead to a simple past reading, when *bì* is combined with a non-stative verb the sentence will have an anterior reading**.** But in this particular context the absence of tense means that the subsequent events all have the same temporal specification as the first event, i.e. an anterior specification. This means that there is a special grammatical rule that applies to narratives only. I propose that this be interpreted as the result of the selection of a specific grammar for narratives, within which the past operator is expressed by means of a zero morpheme in non-initial event descriptions.

A third prominent feature of Saramaccan is the extensive use of serial verb constructions (SVCs), an example of which can be found in (2). According to Veenstra (1996), Saramaccan SVCs are marked for tense only once, which means that they must be considered a single unit. The tense marker occurs before the first verb in the construction and has scope over the construction as a whole. Aspect markers may occur before the second verb, in which case the aspect marker has scope over this verb only.

[7] For a full version of the story I refer to Rountree & Glock (1982)

Usually aspect markers will precede the first verb of a SVC, but as opposed to tense markers, aspect markers may precede the second verb as well. In the latter case these markers will only have a scope over the second verb. This is however a non-preferred and rare construction.

In Foley & Olson's (1985) treatments the verbs in SVCs are part of a single predication, but in the underlying representation they do not form one unit, i.e. the verbs are encoded separately in the grammar. Each verb of a SVC has a meaning of its own, and they can be combined in SVCs to create new meanings. This explains why the verbs that occur in SVCs in Saramaccan can be used as independent main verbs as well. The underlying representation of SVCs should show this, as well as the fact that the two serial verbs share one or more arguments. Both aspects may be taken care of by representing Saramaccan SVCs in the following way (see Hengeveld 1997):

(3) $(\pi_2\ e_1\ [(\pi_1\ f_1)\ (x_1)_\Phi\ (x_2)_\Phi],\ [(\pi_1\ f_2)\ (x_1)_\Phi\ (x_2)_\Phi]\ (e_1))$

Both of the verbs are placed separately between brackets with their own argument slots. Argument sharing is indicated by co-indexing. The verbs and their arguments form part of a single event description (e_1). This representation correctly captures the fact that tense (π_2) has scope over the event as a whole, i.e. over both verbs , while aspect (π_1) may have scope over just the verb. The representation in (3) is used in Figure 2, which provides a full FDG representation of (2).

A final interesting element, which shows up in (1), is the use of free direct speech in (1b). In this case 'direct speech' should not be taken literally, since we are dealing here with an imitation of the sound of the pecking at the rock of Woodpecker: *'kookookoo'*. In Figure 1 I have analysed this sound string as a referent (s_1) at the interpersonal level, since it directly refers to the sound Woodpecker made. The underlying representation is given in (4):

(4) $(R_3: [(s_1: \textit{kookookoo}\ (s_1))]\ (R_3))$

The representation in (4) is used in the representation of (1b) in Figure 1.

3.3. Grammatical particles

After these preliminary issues, let me turn to the actual topic of this paper: grammatical particles in Saramaccan. In the remainder of this paper I will focus on three particles that operate at different levels of analysis: one particle that operates at the level of the discourse act (*hèn* 'then'); one that operates at the level of the referential act (*dí* 'definite'); and one that operates at the level of the event description (*bì* 'past').

The particle *hèn*, illustrated in (1a-c), is a very interesting one. It is very often used by the narrator when a new event in the chronology of the story is presented, except for the first one. Although it could perhaps be analysed as an event marker, since it indicates that the story follows the chronological order of events, I have chosen to interpret is as a discourse marker, since it is used to introduce new steps in a narrative discourse. The folktale is an oral narrative. This implies that, as mentioned earlier, a special grammar is used. In other types of discourse this particle would not be used, or not be used in the same way. This dependency of the marker on the type of discourse is a strong indication that it should be analysed as a discourse marker. In the analysis below the particle *hèn* is therefore inserted in sentence-initial position when a non-initial discourse act within a single narrative is produced. These non-initial discourse acts carry the function Subs(equent), as indicated in (5):

(5) $(M_I: [(A_I), (A_J)_{SUBS} ..... (A_N)_{SUBS}] (M_I))$

(5) represents that move (M_I) consists of several discourse acts (A_I)-(A_J), the non-initial ones carrying the function 'SUBS'.

The particle *dí*, illustrated in (1b), marks definiteness. As such, it represents a relevant distinction at the level of the Referential Act (see Hengeveld forthcoming), which itself is a component of the Interpersonal Level. At this level the speaker estimates the knowledge of the addressee, and one of the possible reflections of this is use of the

particle *dí*, which indicates that the speaker and the addressee share information about the referent. This could be a result of the fact that the referent has been introduced earlier in the discourse, or that it is available within the setting of the speech event. In line (1b) the former situation applies: *sitónu* 'stone' has been introduced by the narrator in the previous part of the narrative. A definite referential act may be represented as in (6):

(6) (d R_1: (x_1))

The presence of the operator 'd' then triggers the insertion of the particle *dí* in the first position of template for referential phrases.

The particle *bì* is the past tense marker in Saramaccan. Whereas in many languages tense is expressed on the verb, Saramaccan uses a separate particle for this category too. The language has a three-way tense system: past tense is expressed by *bì,* future tense by *o*, and the present tense is unmarked. Tense may furthermore remain unmarked once it has been established in a narrative (see above). Only one tense marker is used in an event description, as a result of which only one marker occurs in serial verb constructions. The tense markers are the expression of π_2-operators. The location of these operators was illustrated in (3).

3.4. Analysis

Figure 1 contains the FDG representation of sentence (1b). The relevant particles here are *hèn, dí,* and *Ø*.

The part of the fund that feeds into the formulator (arrow 1) contains frames, lexemes and operators. The interpersonal frame is one that carries the function 'SUBS'. The operators selected are 'd' and 'past'. These three elements have to be mapped onto the particles *hèn, dí,* and a zero-expression. The latter is used for past tense in narrative contexts.

Since the particles are grammatical words rather than morphemes, they have to enter the underlying representation at the morphosyntactic level, where the morphosyntactic encoder first draws the relevant templates from the fund (arrow 8), after which the grammatical particles may be entered into the relevant syntactic slots. Since the zero expression for past occurs in opposition with grammatical particles, it is treated here as a syntactic rather than as a morphological zero-morpheme.

Since all the grammatical distinctions considered are thus expressed by means of particles, the task of the phonological part of the fund is limited to providing prosodic information.

The representation of (2) is given in Figure 2. The relevant particle here is *bì*. The operator 'past' is inserted into the π_2 operator slot within the event frame, and has to be mapped onto the particle *bì*, and since this is a grammatical word, this has to be done by the morphosyntactic encoder again. Note that in this case the clause template contains two predicate phrases, since this is a serial verb construction, and that the particle is expressed as part of the first predicate phrase only.

Figure 1. Analysis of example (1b)

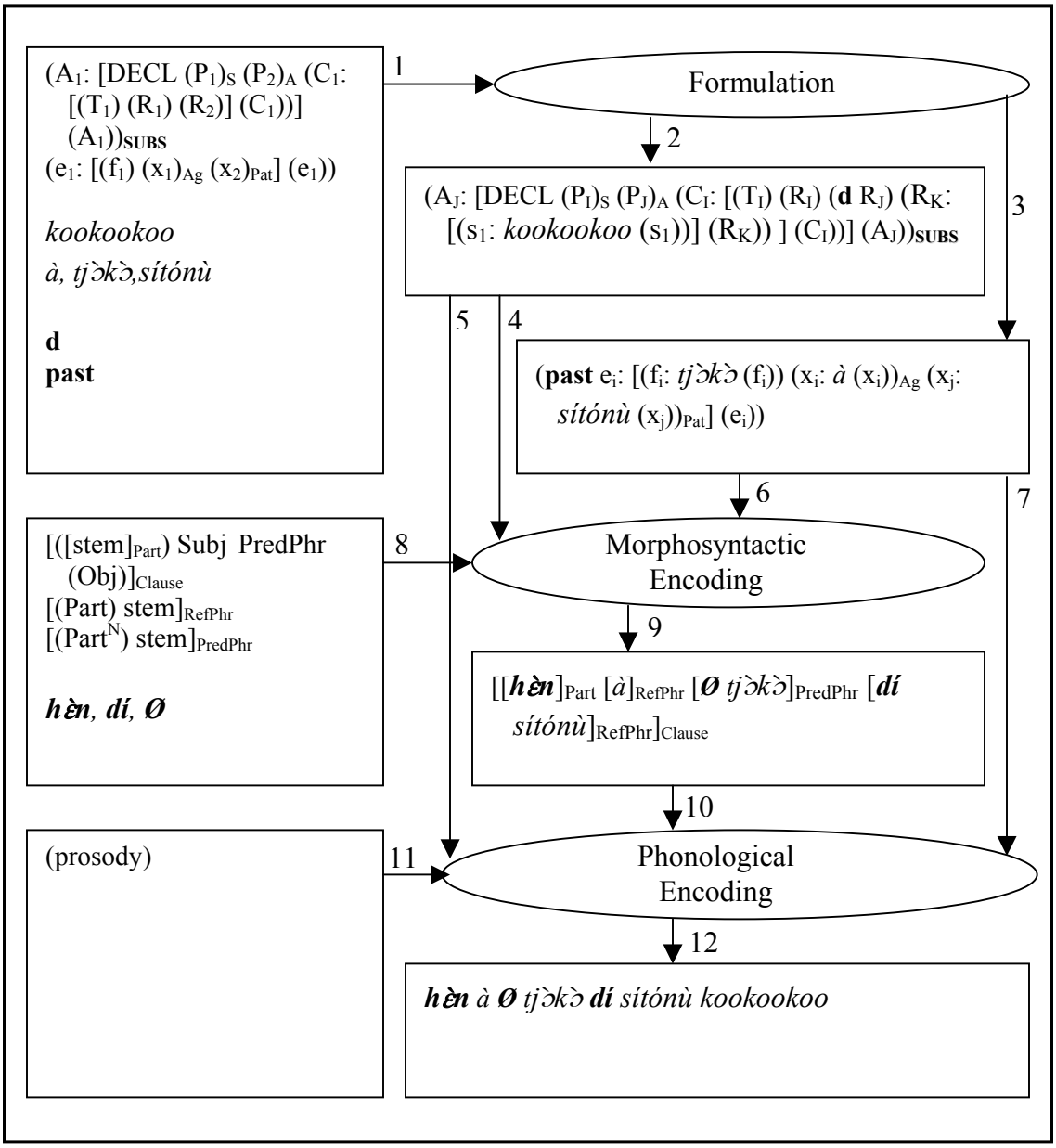
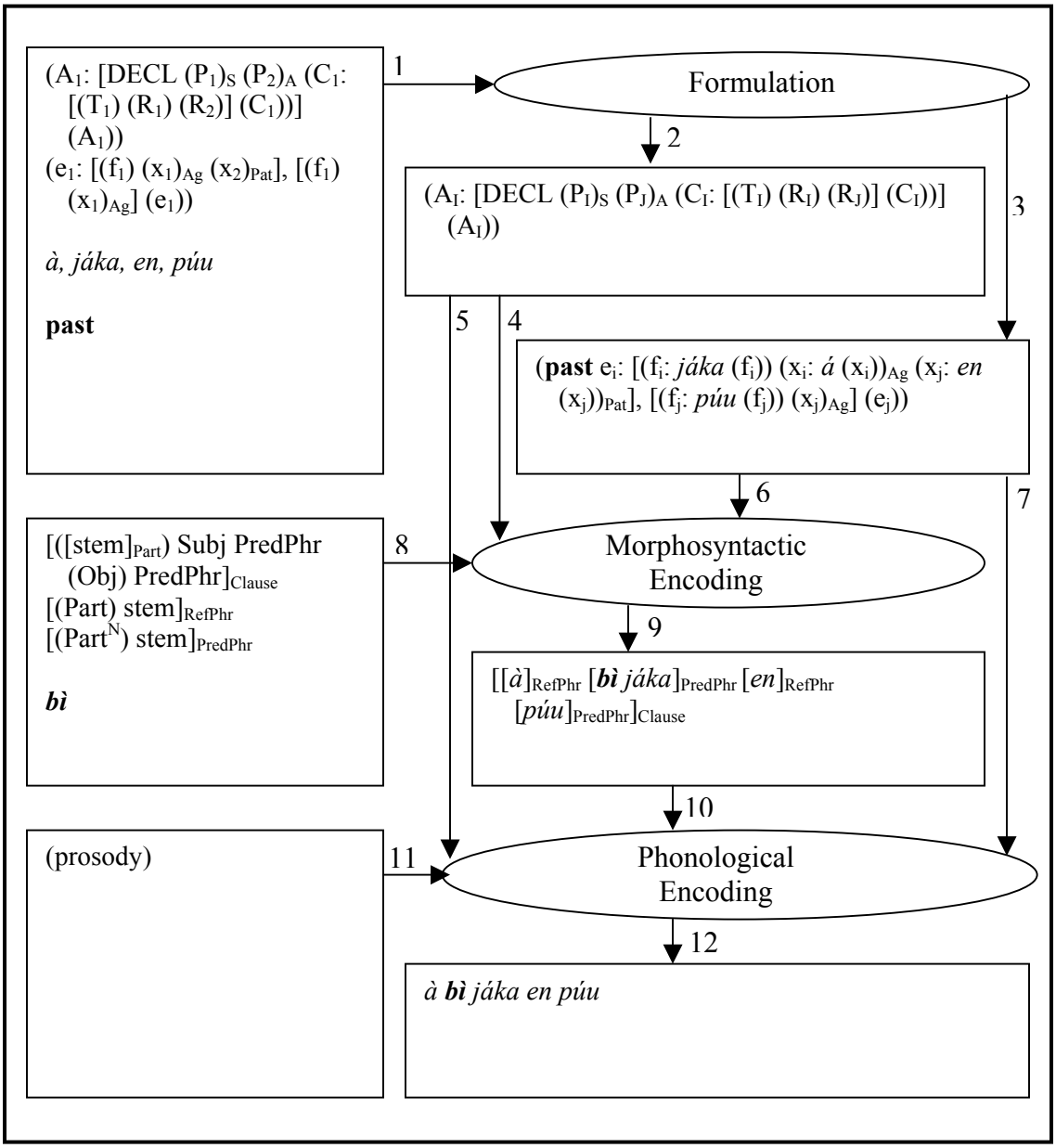


Figure 2. Analysis of example (2)



3.5. Conclusion

It can be deduced from the above representations that FDG is indeed capable of accounting for the wide variety of meanings and functions of grammatical particles at the interpersonal and representational levels, while at the same time accounting for the formal similarity and grammatical behaviour at the morphosyntactic level. The morphosyntactic component provides a variety of templates at the clause and phrase level, into which the grammatical particles, which are stored within the same component, may be inserted. The characteristics of an isolating language are reflected in the fact that all grammatical material is inserted at the morphosyntactic level (rather than at the phonological level, see chapter 2), thus taking away the job of inserting segmental information from the phonological component.

References

Byrne, Francis (1987), *Grammatical relations in a radical creole: Verb complementation in Saramaccan.* Amsterdam: John Benjamins.

Comrie, Bernard (1981), *Language universals and linguistic typology: Syntax and morphology*. Chicago: University of Chicago Press.

Foley, William A. & Michael Olson (1985), Clausehood and verb serialization. In: Johanna Nichols & Anthony C. Woodbury eds, *Grammar inside and outside the clause: some approaches to theory from the field*, 17-60. Cambridge: Cambridge University Press.

Hengeveld, Kees (1997), Serial verb constructions in Functional Grammar. Handout, guest lecture, University of Copenhagen.

Hengeveld, Kees (forthcoming), Dynamic expression in Functional Discourse Grammar. In: Casper de Groot & Kees Hengeveld eds, *Morphosyntactic expression in Functional Grammar* (Functional Grammar Series 27). Berlin: Mouton de Gruyter.

Rountree, S. Catherine & Naomi Glock (1982), *Saramaccan for beginners: a pedagogical grammar of the Saramaccan language* (Languages of the Guianas 5). Paramaribo: Instituut voor Taalwetenschap/Summer Institute of Linguistics.

Veenstra, Tonjes (1996), *Serial verbs in Saramaccan: predication and creole genesis*. The Hague: Holland Academic Graphics.

4. On the whereabouts of gender and number agreement: location and accessibility

Suzanne Dikker

Graduate School for Humanities, University of Amsterdam

4.1. Introduction

The FDG-model as proposed by Hengeveld (2004a, b; see also this volume) constitutes a top-down and strictly modular representation of the generation of linguistic utterances. As Hengeveld (2004b) firmly argues in response to several authors, like Bakker & Siewierska (2004) who adhere to an interpretation of the FDG-model as a processing model, the model does *not* constitute a model of the speaker, though it 'is assumed to be more effective the more closely it resembles this language production process'. One of the problems that arise with the adoption of a processing model of individual speakers rather than a representation of grammar as such is the inevitability of the incorporation of a dynamic flavour (see Bakker 2001 for a proposal for the adoption of Dynamic Expression Rules). Integration of the temporal dimension might force all levels to be linearized, including the representational level.

N-A agreement, the topic of this paper, obviously constitutes a rather dynamic process itself, since it involves the copying of certain features displayed in controllers onto targets (Corbett, 1991). It appears, for example, that adjacency and word-order effects do come into play in agreement, as is shown by Bakker (forthcoming) for verb-subject agreement in Arabic, and can be found in other languages, like Italian, as well. Apart from the impact it has on the morphological output of agreement targets, the copying process itself has been proven to bear particular significance not only in second language acquisition (e.g. Dewaele & Veronique 2001; White et al. 2003), and language impairment (e.g. Bedore & Leonard 2001), but also in language contact situations (e.g. Lipski 1985), and Creoles (Kihm 2002). All these language varieties appear to have in

common that the specific operations responsible for the utterance of agreement are either lacking or are not accessed.

In this discussion of grammatical and semantic gender and number agreement, I focus on the claimed psycholinguistic validity of the FDG model, by relating the results of psycholinguistic studies on gender and number agreement to the FDG model. Findings in the fields of Psycholinguistics and Neurolinguistics might give insight into the level of in(ter)dependence and interrelatedness of the respective grammar modules. Since the concept of gender (and number) is relevant at different levels, it forms an attractive testing ground for the strict modular and top-down character of the FDG-model.

4.2. Conceptual versus grammatical information

A distinction has to be made between semantic/conceptual gender on the one hand, and grammatical gender on the other. The former reflects biological gender, while the latter does not. That there is no one-to-one relationship between the two is exemplified by a word like *persona* in Spanish and Italian, which is inherently feminine in terms of grammatical gender, but which can refer to both men and women. Semantic gender is most logically drawn from the Contextual Component, since its expression depends on the biological gender of external referents relevant to the ongoing discourse. Grammatical gender, on the other hand, is an inherent feature of the lexeme involved, and therefore coded as a syntactic feature in the Fund. Consequently, grammatical gender is triggered within the grammar, while semantic gender is triggered by extra-grammatical information.

Basically the same can be said with respect to conceptual number versus grammatical number. Supporting the current representation of number and gender within the FDG model, results from several psycholinguistic studies suggest that gender and number features are used at different processing stages (De Vincenzi, 1999; Dominguez et al, 1999). Nevertheless, just as with gender, number information can be grammatical as well as notional in nature, and thereby possibly in conflict. Consider for example the Spanish word *gente* 'people', which is inherently singular and feminine, but generally carries a plural and neuter interpretation in extra-linguistic reference.

There are basically two proposals in the psycholinguistic literature concerning the relation between conceptual and grammatical gender and number (Vigliocco & Franck, 1999). The **Minimal Input Hypothesis** predicts that agreement is a purely syntactic operation, involving the copying of syntactic features of controllers onto targets. As soon as these features are retrieved, information from the conceptual structure, i.e. semantic considerations, is no longer of influence on the output. The **Maximal Input Hypothesis,** on the other hand, predicts that conceptual information that can be of influence on the agreement features is accessible throughout the whole process of sentence generation. This hypothesis follows from psycholinguistic models in which "feature-merging" functions as a mechanism for the building of syntactic structures. In support of this hypothesis, Vigliocco & Frank (1999: 458) give examples of so-called Notional Concord found in spontaneous speech, as in (1) (Italian):

(1)	*Siamo*	*un-a*	*squadra*	*che*	*gioch-iamo*
	be.PRS.**1PL**	INDEF-F.**SG**	team-**SG**	COMP	play-PRS.**1PL**

'We are a team that play .'

Similarly, they point out that there are numerous examples in speech in which semantic gender turns out to be more prominent than grammatical gender. Returning to the noun *persona*, which was used above to exemplify the difference between conceptual and grammatical gender, consider example (2), taken from Vigliocco & Frank (1999: 456), in which the semantic gender of the referents (masculine) overrules the grammatical gender of the controller (feminine).

(2)	*Sono*	*sces-e*		*quattro*	*persone,*
	be.PRS.3PL	get.off.PTCP-**F.PL**		four	person.**PL(F)**
	de-i	*qual-i*	*tre*	*sono*	*andat-i.*
	POSS-**M.PL**	which-PL	three	be.PRS.3PL	go.PTCP-**M.PL**

'Four persons got off; of the four, three went.'

Vigliocco & Franck (1999) further state that on a general note, fewer errors occur when agreement is based on or strengthened by conceptual gender, than in cases of pure syntactic agreement.

4.3. Theoretical implications

How do these two hypotheses, together with the empirical findings, relate to the FDG model? The FDG model as proposed here clearly follows the Minimal Input Hypothesis. Conceptual gender information is provided by the Contextual Component, and since the Contextual Component only feeds into the Formulation level, one would have to assume that cases of notional concord like example (2) above, are conveyed in two steps: in the first step, the part *sono (...) persone* is generated. Then, the speaker returns to the top module again, where the contextual information can be re-accessed, in this case with respect to the semantic gender of the referent. This sentence in fact does not form a great threat to the FDG model as formulated above, since one could easily argue that it consists of two separate Discourse Acts, thereby motivating the return to the top of the model. Sentences like (1) on the other hand, with a restrictive rather than a nonrestrictive relative clause, do pose a problem, since there one cannot argue that going back to the Interpersonal Level is motivated by the formation of a new Discourse Act. The same goes for verb-subject agreement. Consider examples (3a-b), adapted from Vigliocco & Frank (1999: 458):

(3) a. ****These*** *committee voted for a raise*

b. *The committee* ***are*** *voting themselves a raise*

The difference between (3a) and (3b) can be accounted for in a similar fashion as the Italian example above, i.e. in terms of processing units: in (3a) the demonstrative is contained within the same template, namely that of the referential phrase, and is thereby subject to purely formal considerations and blind to semantics. In (3b) the generation process, after expressing the term and its subordinate node, continues by returning to the level at which the contextual component can be accessed.

The unit-generation approach is comparable to ideas of parallel processing. But then the FDG-model would have to be viewed as a model of the speaker, i.e. a processing model (Bakker & Siewierska 2004), in contrast to the competence representation of grammar adopted here. Levels and limits of access are relevant to a static model as well, though, since examples like (3b) are fully grammaticalized. Consider the Agreement Hierarchy proposed in Corbett (1991: 226):

(4) Attributive < predicate < relative pronoun < personal pronoun

As Corbett (1991:226) states: 'As we move rightwards along the hierarchy, the likelihood of semantic agreement will increase monotonically (that is, with no intervening decrease)'. This hierarchy quite neatly reflects the examples given above (assuming that semantic interference is relevant for the consideration of number as well as gender), and thereby supports a unit-based model. These units would then be of a local character, and the accessibility of grammatical gender features should be defined in local terms as well: the more local borders are crossed, the greater the chance for semantic information to take the lead over syntactic information. One could include determiners under the attributive agreement targets, and subdivide the category of relative pronoun into restrictive and non-restrictive relative pronouns, a division that would form the frontier between targets within a single Discourse Act and agreement that crosses the border between two discourse units. Let me elaborate a little bit on the concept of locality and its influence on the utterance of gender agreement.

A relevant question to the present topic is the representation of gender and number. Semantic gender and number information was said to be drawn from the Contextual Component, and therefore does not need to be accounted for at the level of the grammar. As for grammatical gender and number, at first sight, the most logical place for them to be represented would be at the Representational Level, since they constitute an inherent feature of the lemma. With regard to grammatical gender and number, representations like the one in (5) seem attractive:

(5) $(1\ x_i: \text{squadra}_{\text{NfemNpl}}(x_1))$ → *una squadra* ('a team')

Schriefers & Jescheniak (1999), however, conclude from a large number of studies on the retrieval of grammatical gender that the selection of a lexical element does not necessarily entail the selection of all of its inherent features: grammatical gender is only triggered if it is needed in its local syntactic environment, i.e. for agreement purposes. Consequently, it would depend on the desirability of coding "unused" properties in the UR whether or not grammatical gender should be represented. Since it is apparently not triggered if there is no target close by, for reasons of economy I would argue that in those cases gender is not taken up in the UR.

An implication of such an analysis is that a heavy burden is put on the nature and presence of agreement targets, since they "decide" whether gender is to be triggered or not. Secondly, one would need to define the locality of syntactic environment, and set it in contrast to the occurrence of gender agreement across the frontiers of this locus. From the hierarchy formulated above, it seems that the cluster of a head of a Referential Phrase and its attributive adjective(s) and determiners can be defined as the local syntactic domain of a grammatical gender/number agreement trigger. Note that relative clauses appear after predicates in the hierarchy. This suggests that the barrier between heads and relative clauses is stronger than the one between triggers and predicative targets. This should have a correlate in the order in which the elements of an utterance are implemented into the templates provided at the level of morphosyntactic encoding. Back to the relation between local domains and agreement: would this mean that - if the local environment has not triggered the retrieval of the grammatical gender of the noun (bare nouns) - one would need to re-retrieve the lexeme of the controller from the Fund if it turns out that its gender features are required at a later point in the utterance? These questions all boil down to what was already pointed out by Bakker & Siewierska (2004: 357), i.e. the question of '[…] the way information is stored [in the Fund], what is retrievable, […], and under what conditions'. For the grammatical gender features of a lemma it thus appears that they are retrievable under the condition that there is a target nearby. ERP[8] data further suggest that retrieving the gender information belonging to a lemma is an extra step after lemma access is achieved, and appears to occur parallel to

[8] Event-Related Brain Potential

the retrieval of the word form of the lexeme under consideration (Schmitt et al 2000; 2001). It is important to stress that the observation that 'if a language has targets which distinguish gender, then typically they *must* distinguish genders' (Corbett, 1991:218) suggests a constant driving force behind the retrieval of gender information.

The FDG-model is defined as strictly modular and top-down, and the assumption that gender information is coded on the lexeme has the consequence that it can only be retrieved at the level of Formulation, at least within a representation of the Fund as proposed. The restriction of this retrieval was defined above in terms of necessity, i.e. if the local domain requires it. Although there is an obvious need for a more detailed discussion, here I would like to roughly define the boundaries of the local syntactic environment in terms of the agreement hierarchy defined above, including the slight refinements suggested.

Integrating these observations leads to the following two hypotheses: (i) languages are more likely to show gender agreement within the local syntactic environment of the trigger than outside it, and (ii) if languages show agreement on targets outside the local syntactic environment, they prefer to do so in combination with the expression of agreement targets within local syntactic government, since otherwise one is forced to go through all the levels of the grammar only to retrieve the grammatical gender of a noun. The first hypothesis can be related to Corbett's (1991: 1) observation that 'elements which are morphologically capable of showing gender agreement may be restricted to doing so in specific syntactic environments'. As an example he mentions German, which shows gender agreement on the adjective in attributive position, but not in predicative position. Spanish, on the other hand, does portray gender agreement on predicative targets. Based on the second hypothesis one would then expect that Spanish does not allow bare nouns in term position in these cases, since this would imply that the lexical information related to the controller needs to be accessed twice. Indeed, the use of bare nominals is much more restricted in Spanish (7) than it is in Dutch (6), as is shown in the generic use of nouns (Leonetti 1999):

(6) *Nederlander-s* *zijn* *dom.*
Dutch. person-PL COP.PL stupid
'The Dutch are stupid.'

(7) a *Los* *holandes-es* *son* *tonto-s.*
the Dutch. person-PL COP.3.PL stupid-PL
'The Dutch are stupid.'

b **Holandeses* *son* *tonto-s.*
Dutch. person-PL COP.3.PL stupid-PL
'The Dutch are stupid.'

This would, of course, only apply to languages for which it can be argued that determiners are agreement targets for grammatical gender. Needless to say, examples from two Indo-European languages do not suffice to validate a hypothesis, but the relation between syntactic scope and syntactic gender agreement is in fact reflected in the Agreement Hierarchy (Corbett 1991: 226)

If typological data turn out to support the hypotheses mentioned above, the value of a strictly modular and top-down model is supported in a neat way: instead of having to resort to the formulation of all kinds of restrictions and barriers to explain the limits of gender assignment, general theoretical principles suffice. But what if the data do *not* support the hypotheses? This would have as a consequence that in those languages that regularly show gender agreement on targets outside the local syntactic environment of the controller without the presence of a trigger in the local domain, the generation of gender agreement morphology becomes a very “expensive” operation, since one keeps on having to go back to the Formulator only to retrieve a syntactic feature coded on the lexeme from the Fund. Then these theoretical principles all of a sudden become a problematic factor, and one would prefer to have constant access to the Fund for economy purposes.

4.4. Levels of Access and Locality

I will try to find a solution to the above-mentioned problems within the model as proposed in chapter 1, i.e. having a strictly top-down and modular character. The first question to be addressed is if one indeed needs to resort back to a higher level after the consideration of a local chunk, to which level one would have to return.

Above, it was assumed that grammatical gender, if relevant, is coded in the UR. One could ask, however, if grammatical gender is ever relevant at the RL, since apparently it does not come into use until the local syntactic domain is considered, i.e. at the level of morphosyntactic encoding. Adopting an analysis in which grammatical gender is retrieved from the Fund during Morphosyntactic Encoding, and not during Formulation, has the advantage that one does not need to return to the very top of the model if syntactic features of the lexemes are to be retrieved. After the formulation of the Representational Level, the main sentence template is created in the Morphosyntactic Encoder, on the basis of combined information from both the RL and the IL. From there on, each slot, consisting of various sub-nodes, is provided with morphosyntactic information retrieved from the Fund. These retrieved features are inherited by subnodes, and are accessible to higher nodes by means of percolation (see Bakker 2001 for a discussion of the principles involved in the Dynamic Expression Rules).

I would like to argue that both locally bound features and global features are relevant to the output of the structural level. The latter are fed into the Morphosyntactic Encoder and accessible at all times, while the first are fed from the Fund into the restricted local environment (i.e. one of the main nodes of the sentence template and its subordinate nodes). The locally bound (formal) features can be subject to barriers (see Bakker 2001 and Bakker & Siewierska 2002), while the global features cannot. Grammatical gender and number would be locally bound features, since they are presumably to be retrieved from the Fund only when relevant to the local syntactic environment. Conceptual gender and number would be global features. This means that the Contextual Component has to feed into the Morphosyntactic Encoder directly.

At each local level of morphosyntactic encoding, locally bound formal features have priority over other information from the Morphosyntactic Encoder. This explains

the ungrammaticality of (3)a. As soon as the structural level of a local environment is formulated, some features are passed on by inheritance to the next main slot to be filled with morpho-syntactic information. In some languages features like grammatical gender are not taken to the next slot, while in other languages they are transferred, but enter into competition with global features, as was shown for English in (3b) and Italian in (2) above.

I have formalized these ideas in Figure 1 below, which consists of a slightly adapted version of the FDG-model as formulated in the chapter 1. The adaptations consist of: (i) the specification of the content of the level of Morphosyntactic Encoding in terms of definition of the relevant templates (P1, P2 and P3); (ii) the inclusion of the locally bound features, which are retrieved from the relevant lexeme through the second box of the Fund (as secondary operators) and fed into the local units of morphosyntactic encoding (MS1, MS2, and MS3); and (iii) an arrow feeding into the Morphosyntactic Encoder from the Conceptual Component, providing it with the global features (like conceptual gender). The representation in Figure 1 incorporates the following adaptations to the initial FDG-model:

(i) Morphosyntactic encoding (MS) is dynamicized internally, in the sense that after the retrieval of the relevant templates, these are encoded subsequently:

P1→ MS1 → 1 / P2 → MS2 → 2 / P3 → MS3 → 3

At the beginning of this discussion, I pointed out that the introduction of dynamicity could force the undesirable situation that all levels are to be linearized, including the RL. In the current proposal, linearization is postponed until the level at which linear order is strictly relevant, i.e. with the insertion of Templates. I would like to refer to Bakker's (2001) elaboration of the Dynamic Expression Rules for a specification of the possible principles that guide morphosyntactic encoding (top-down, left-right, depth-first, inheritance and percolation). For grammatical gender and number agreement, this dynamicity is motivated by the consideration of local domains: it only makes sense to talk about local agreement domains if syntactic information is encoded bit by bit. MS1, MS2 and MS3 each represent a separate domain here.

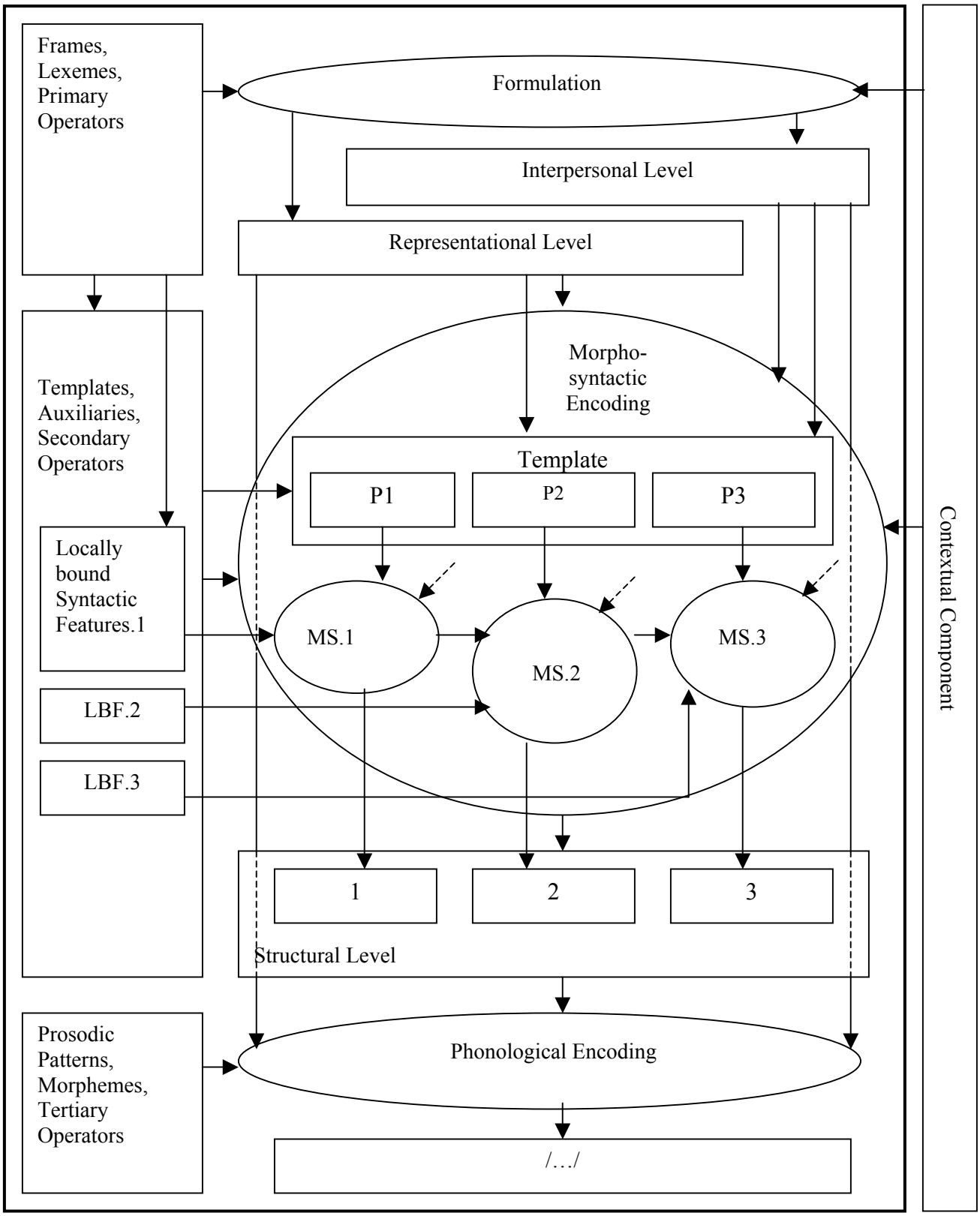
(ii) The first and the second box of the Fund are linked: the second box contains the syntactic features of the lexemes specified in the first box, defined in the representation above as Locally Bound Syntactic Features, like the grammatical gender and number information needed for agreement purposes. The triggering of these features is a real-time operation, which occurs simultaneously to the sending of information from the representational level to the Morphosyntactic Encoder. This accounts for the ERP data that suggest this parallel processing, as pointed out above (Schmitt et al 2000; 2001).

(iii) The grammatical gender and number information contained in the Locally Bound Features LBF1, LBF2, and LBF3 is sent to the local syntactic domains that are being encoded at MS1, MS2, and MS3 respectively. At this point, the LBFs have exclusive access to their corresponding MSs, and thereby overrule other possible information from the rest of the Morphosyntactic Encoder. The arrows that seem to appear out of nowhere into the MSs are meant to represent the access of these other types of information. This representation would account for the ungrammaticality of sentences like (3a). In SLA, specific language impairment and Creole formation, it might be the case that the LBFs are not accessed, as a result of which grammatical agreement fails. One would expect conceptual information to provide the repair kit for this lack. It is indeed the case that some Creole languages, like Angolar (Lorenzino 1998), show semantic agreement in some cases, but no grammatical agreement at all. Indirect evidence from sentence processing is provided by the observation that right after the input of a grammatical violation, semantic repair resources are called upon (ERP data reported in e.g. Hagoort et al. 1993).

(iv) LBF1 can be passed on by inheritance or percolated to MS2, and LBF2 to MS3, but in this process these features lose their exclusivity: other information contained in the Morphosyntactic Encoder, for example conceptual gender and number, now competes in strength with the LBFs. This accounts for the occurrence of sentences like (3b) and (1).

(v) In order for conceptual information to be economically accessible to the Morphosyntactic Encoder, i.e. to avoid the necessity of returning to higher levels to retrieve this information, in the representation given above it is proposed that the Conceptual Component feeds into the Morphosyntactic Encoder directly.

Figure 1. An adapted version of the FDG model



The last problem to be taken into consideration in the realization of grammatical gender agreement is the fact that it is only triggered by the presence of a target in the local environment of the controller (i.e. when in attributive position). The prediction remains that languages that portray grammatical gender agreement outside the borders of the local domain of the controller will show a preference against the use of bare nouns (see 7 above for Spanish). The question how one would re-access the locally bound features still remains valid, however, since preference entails non-absoluteness. It would have to be possible, though less attractive. Consider the following hypothetical example, concerning a language which shows predicative gender agreement, and allows for bare nouns: P1 contains a bare noun, and therefore its grammatical gender is not retrieved at MS1. P3, however, contains a predicative adjective, and thereby needs the information of the grammatical gender of the noun expressed in 1 in order to yield the appropriate morphological expression. One could assume that some kind of priming effect facilitates the retrieval of the relevant LBFs: The pathway to the lexeme has been activated at MS1. At MS3, this pathway is still more active than others, due to recency effects, and consequently its related features can be accessed pretty easily from the Fund directly, by calling upon memory.

4.5. Conclusion

I by no means pretend to have provided all the answers to the issues involved in grammatical and semantic agreement, and flaws and inaccuracies will undoubtedly be discovered shortly, but I do hope to have pointed out which problems the FDG model as proposed in chapter 1 faces when it needs to account for some of the agreement phenomena that are observed by typologists, psycholinguists and neurolinguists.

References

Bakker, Dik (2001), The FG expression rules: A dynamic model. *Revista Canaria de Estudios Ingleses* 42, 15-53.

Bakker, Dik (forthcoming), Agreement: more arguments for the dynamic expression model. In: Casper de Groot & Kees Hengeveld eds, *Morphosyntactic expression in Functional Grammar* (Functional Grammar Series 27). Berlin: Mouton de Gruyter.

Bakker, Dik & Anna Siewierska (2002), Adpositions, the lexicon and expression rules. In: R. Mairal Usón & M.J. Pérez Quintero eds, *New perspectives on argument structure in Functional Grammar.* Berlin: Mouton de Gruyter, 125-178.

Bakker, Dik & Anna Siewierska (2004), Towards a speaker model of Functional Grammar. In: J. Lachlan Mackenzie & María Á. Gómez González (eds), *A new architecture for Functional Grammar* (Functional Grammar Series 24), 325-364. Berlin: Mouton de Gruyter.

Bedore, Lisa & Laurence B. Leonard (2001), Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of Speech, Language and Hearing Research* 44, 886-904.

Corbett, Greville (1991), *Gender.* Cambridge: Cambridge University Press.

Dewaele, Jean-Marc & Daniel Véronique (2001), Gender assignment and gender agreement in advanced French interlanguage: A cross-sectional study. *Bilingualism: Language and Cognition* 4.3, 275-297.

Dominguez, Alberto, Juan Segui & Fernando Cuetos (1999), The processing of grammatical gender and number in Spanish. *Journal of Psycholinguistics* 28.5, 485-498.

Hagoort, Peter, Colin Brown & Jolanda Groothusen (1993), The syntactic positive shift (SPS) as an ERP measure of syntactic processing. *Languages and Cognitive Processes* 8.4, 439-483.

Hengeveld, Kees (2004a), The architecture of a Functional Discourse Grammar. In: J. Lachlan Mackenzie & María Á. Gómez González (eds), *A new architecture for Functional Grammar* (Functional Grammar Series 24). Berlin: Mouton de Gruyter, 1-21.

Hengeveld, Kees (2004b), Epilogue. In: J. Lachlan Mackenzie & María Á. Gómez González (eds), *A new architecture for Functional Grammar* (Functional Grammar Series 24). Berlin: Mouton de Gruyter, 365-378.

Kihm, Alain (2003), Inflectional categories in creole languages. In : Ingo Plag (ed.) *Phonology and morphology of creole languages*. Tubingen: Niemeyer, 333-336.

Leonetti Jungl, Manuel (1999), El artículo. In: Ignacio Bosque & Violeta Demonte (eds.), *Gramática descriptiva de la lengua española,*Vol. I. Madrid: Espasa, 787-890.

Lipski, John (1985), *The Spanish of Equatorial Guinea: The dialect of Malabo and the Implications for Spanish dialectology.* Tübingen: Max Niemeyer.

Lorenzino, Gerardo A. (2000), *The morphosyntax of Spanish-lexified creoles.* München: Lincom Europa.

Schriefers, H. & J.D. Jescheniak (1999), Representation and processing of grammatical gender in language production. *Journal of Psycholinguistics* 28.6, 575-600.

Schmitt, Bernadette M. et al (2000), Electrophysiological estimates of the time course of semantic and phonological encoding during implicit picture naming. *Psychophysiology* 37.4, 473-484.

Schmitt, Bernadette M. et al (2001), An electrophysiological analysis of the time course of conceptual and syntactic encoding during tacit picture naming. *Journal of Cognitive Neuroscience* 13.4, 510-522.

De Vincenzi, Marica (1999), Differences between the morphology of gender and number: Evidence from establishing coreferences. *Journal of Psycholinguistics* 28.5, 537-553.

Vigliocco, Gabriella & Julie Franck (1999), When sex and syntax go hand in hand: Gender and agreement in language production. *Journal of Memory and Language* 40, 455-478.

White, Lydia et al (2003), Gender agreement in L2 Spanish: Evidence against failed features. *McGill Working Papers in Linguistics* 17.2, 153-182.