Parts-of-speech system as a basic typological determinant

Kees Hengeveld

1 Introduction

In Hengeveld (1992a, 1992b) I propose a functional theory of parts of speech (PoS) that posits a fundamental distinction between flexible and rigid PoS-systems: flexible PoS-systems display classes of lexemes that can be used in more than one function without requiring lexical or syntactic derivation, while rigid PoS-systems display classes of lexemes that are tied to a single function and require lexical or syntactic derivation in order to be used in other functions. Further differentiation between PoS-systems is shown to be due to an implicational hierarchy that systematically defines different degrees of flexibility and rigidity in PoS-systems.

Later work by various authors has shown that the resulting typology of PoS-systems, and especially the major split between flexible and rigid languages, leads to strong predictions about the organization of the syntax, morphology, and lexicon of languages with different types of PoS-systems, and especially of those with a flexible PoS-system. It is the aim of this paper to bring these results together so as to come to an overall assessment of the impact of PoS-systems on the grammar of languages. The paper thus claims no originality as to the individual phenomena reported on, but tries to provide an integrated view of these. For reasons of space the various studies reported on can each be presented and exemplified only briefly. For details I refer the reader to the original papers. The language samples on which these studies are based show considerable overlap but also differ in size and composition. Appendix 1 provides a tabular overview of the samples used in the various studies referred to.

After briefly summarizing the parts-of-speech theory proposed in Hengeveld (1992a, 1992b) and refined in Hengeveld, Rijkhoff and Siewierska (2004) in section 2, I introduce, in section 3, the predictions that follow from it with respect to other properties of the languages involved. The predictions are presented in four groups, which have to do with the identifiability of constituents (section 4), the formal integrity of lexemes (section 5), the morphological and semantic unity of classes of lexemes (section 6), and the pervasiveness of flexibility or rigidity in the grammar as a whole (section 7). Section 8 then brings together the various results and gives a general characterization of types of languages in terms of combinations of lexical, morphological, and syntactic features.

2 Flexibility, rigidity, and the parts-of-speech hierarchy[1]

Hengeveld (1992a, 1992b) classifies basic and derived lexemes in terms of their distribution across the four functional slots given in Figure 1.

	HEAD	MODIFIER
PREDICATE PHRASE	V	MAdv
REFERENTIAL PHRASE	N	Adj

Fig.1. Lexemes and functions

[1] This section is largely based on earlier summaries of the model, such as the ones in Hengeveld (2007) and Hengeveld and van Lier (2008, 2010).

Figure 1 shows that the four functional positions are based on two parameters, one involving the opposition between predication and reference, the other between heads and modifiers. Together, these two parameters define the following four functions: head and modifier of a predicate phrase, and head and modifier of a referential phrase. The four functions and their lexical expression can be illustrated by means of the English sentence in (1).

(1) *The tall*$_{\text{A}}$ *girl*$_{\text{N}}$ *sings*$_{\text{V}}$ *beautifully*$_{\text{MAdv}}$

English can be said to display separate lexeme classes of verbs, nouns, adjectives and (derived) manner adverbs, on the basis of the distribution of these classes across the four functions identified in Figure 1: verbs like *sing* are used as heads of predicate phrases; nouns like *girl* as heads of referential phrases; adjectives like *tall* as modifiers in referential phrases; and manner adverbs like *beautifully* as modifiers in predicate phrases. Thus, in this example there is a one-to-one relation between function and lexeme class. Parts-of-speech systems of this type are called differentiated, as for each function there is a separate class of lexemes.

The four categories of lexemes in Figure 1 may be defined as follows: a verb (V) is a lexeme that can be used as the head of a predicate phrase only; a noun (N) is a lexeme that can be used as the head of a referential phrase; an adjective (A) is a lexeme that can be used as a modifier within a referential phrase; and a manner adverb (MAdv) is a lexeme that can be used as a modifier within a predicate phrase. Note that within the class of adverbs I restrict myself to manner adverbs. I exclude other classes of adverbs, such as temporal and spatial ones, since these do not modify the head of the predicate phrase, but rather modify the sentence as a whole. The restriction imposed on verbs that they can be used predicatively *only* is not paralleled in the definitions of the other lexeme classes, as these in many languages allow a predicative use apart from their distinguishing non-predicative functions.

There are other parts-of-speech systems in which there is no one-to-one relation between the four functions identified and the lexeme classes available. These systems are of two types. In the first type, a single class of lexemes is used in more than one function. Such lexeme classes, and the parts-of-speech systems in which they appear, are called *flexible*. The second type is called *rigid*. Rigid systems resemble differentiated systems to the extent that both consist only of lexemes classes that are specialized, i.e. dedicated to the expression of a single function. However, rigid systems are characterized by the fact that they do not have four lexeme classes, one for each of the four functions. Rather, for one or more functions a dedicated lexeme class is lacking. The following examples illustrate the difference between these flexible and rigid parts-of-speech systems. In Turkish (Göksel and Kerslake 2005: 49) the same lexical item may be used indiscriminately as the head of a referential phrase (2), as a modifier within a referential phrase (3), and as a modifier within a predicate phrase (4):

(2) *güzel-im*
beauty-1.POSS
'my beauty'

(3) *güzel bir köpek*
beauty ART dog
'a beautiful dog'

(4) *Güzel konuş-tu-Ø*
beauty speak-PST-3.SG
'S/he spoke well.'

The situation in Krongo is rather different. This language has basic classes of nouns and verbs, but not of adjectives and manner adverbs. In order to modify a head noun within a referential phrase, a relative clause has to be formed on the basis of a verbal lexeme, as illustrated in (5) and (6) (Reh 1985: 251):

(5) *Álímì bìitì*
be.cold.M.IPFV water
'The water is cold.'

(6) *bìitì ŋ-álímì*
water CONN-be.cold.M.IPFV
'cold water' (lit. 'water that is cold')

In (6) the inflected verb form *álímì* 'is cold' is used within a relative clause introduced by the bound connective *ŋ-* or one of its allomorphs. This is the general relativizing strategy in Krongo, as illustrated by the following examples (Reh 1985: 256):

(7) *N-úllà àʔàŋ kí-ǹt-àndiŋ n-úufɔ̀-ŋ kɔ̀-nìimɔ̀ kàti*
1/2-love.IPFV I LOC-SG-clothes CONN:N-sew.IPFV-TR POSS-mother my
'I love the dress that my mother is sewing.'

(8) *káaw m-àasàlàa-tí àakù*
person CONN:F-look.PFV-1.SG she
'the woman that I looked at (her)'

This shows that *álímì* in (6) is not a lexically derived adjective but a verb that serves as the main predicate of a relative clause. Since this is the only attributive strategy available in Krongo, one may conclude that the function of adnominal modification is expressed by relative clauses in this language, not by lexical modifiers.

The same strategy is used to modify a verbal head within a predicate phrase, as illustrated in (9) (Reh 1985: 345):

(9) *ŋ-áa árící ádìyà kítáccì-mày ɲ-íisò túkkúrú.kúbú*
CONN.M-COP man come.INF there-REF **CONN**.M.IPFV-walk with.low.head
'The man arrived walking with his head down.'

The bound subordinating connector morpheme is added to the verb form *íisò* 'walk' in (9). This verb again fulfils the function of head of a predicate phrase within the adverbial subordinate clause, which as a whole fulfils the function of modifier in the (main) predicate phrase.

In sum, the difference between English (differentiated), Turkish (flexible), and Krongo (rigid), is thus that (i) Turkish has a class of flexible lexical items that may be used in several functions, where English uses three specialized classes (nouns, adjectives, and manner adverbs), and that (ii) Krongo lacks classes of lexical items for the modifier functions, where English does have lexical classes of adjectives and manner adverbs.

Krongo has to resort to alternative syntactic strategies to compensate for the absence of a lexical solution. These differences may be represented as in Fig. 2.

language	*head of pred. phrase*	*head of ref. phrase*	*modifier of ref. phrase*	*modifier of pred. phrase*
Turkish	verb	non-verb		
English	verb	noun	adjective	manner adverb
Krongo	verb	noun	-	-

Fig. 2: Flexible, differentiated, and rigid languages

As Figure 2 shows, Turkish and Krongo are similar in that they have two main classes of lexemes. They are radically different, however, in the extent to which one of these classes may be used in the construction of predications: the Turkish class of non-verbs may be used in three functions, while the Krongo class of nouns may be used as the head of a referential phrase only. Notice that for a lexeme class to be classified as flexible, the flexibility should not be a property of a subset of items, but a general feature of the entire class.

Hengeveld (1992a, 1992b) and Hengeveld, Rijkhoff and Siewierska (2004) argue that the arrangement of the functions in Figure 2 is not a coincidence. It is claimed to reflect the parts-of-speech hierarchy in (10):

(10) Head of Pred. phrase > Head of Ref. phrase > Modifier of Ref. phrase > Modifier of Pred. phrase

The more to the left a function is on this hierarchy, the more likely it is that a language has a specialized class of lexemes to express that function and the more to the right, the less likely. The hierarchy is implicational, so that, for example, if a language has a specialized class of lexemes to fulfil the function of modifier of a referential phrase, i.e. adjectives, then it will also have specialized classes of lexemes for the functions of head of a referential phrase, i.e. nouns, and head of a predicate phrase, i.e. verbs. In addition, if a language has a flexible lexeme class that can be used to express the functions of head of a referential phrase and modifier in a predicate phrase, then it is predicted that this class can also be used for the expression of the function lying in between these two in the hierarchy, namely modifier in a referential phrase. Similarly, if a language has no lexeme class for the function of modifier in a referential phrase (i.e. no adjectives), it will neither have a lexeme class for the function of modifier in a predicate phrase (i.e. manner adverbs). Note that the hierarchy makes no claims about adverbs other than those of manner.

The hierarchy in (10), combined with the distinction between flexible, differentiated, and rigid languages, predicts a set of seven possible parts-of-speech systems, which is represented in Figure 3. As this figure shows, it is predicted that languages can display three different degrees of flexibility (systems 1-3), three different degrees of rigidity (systems 5-7), or can be differentiated (type 4). Of the languages discussed earlier Turkish would be a type 2 language, English a type 4 language, and Krongo a type 6 language. Note that I use the term ‘contentive’ for lexical elements that may appear in any of the four functions distinguished. The term 'modifier' is used for lexemes that may be used as modifiers in both predicative and referential phrases.

PoS system		*head of pred. phrase*	*head of ref. phrase*	*modifier of ref. phrase*	*modifier of pred. phrase*
Flexible	1	contentive			
	2	verb	non-verb		
	3	verb	noun	modifier	
Differentiated	4	verb	noun	adjective	manner adverb
Rigid	5	verb	noun	adjective	
	6	verb	noun		
	7	verb			

Figure 3: Parts-of-speech systems

In addition to the seven types listed in Figure 3, there are so-called intermediate systems, showing characteristics of two systems that are contiguous in Figure 3. In flexible languages the most common source for such an intermediate status is that derived stems show a lower degree of flexibility than basic stems. In rigid languages the most common source of an intermediate status is the existence of small, closed classes of lexemes at the fringe of the system. Figure 4 (see also Smit 2007) shows the full set of possible systems, including the intermediate ones.

		head of pred.phrase	*head of ref. phrase*	*modifier of ref. phrase*	*modifier of pred. phrase*
Flexible	1	contentive			
	1/2	contentive			
			non-verb		
	2	verb	non-verb		
	2/3	verb	non-verb		
				modifier	
	3	verb	noun	modifier	
	3/4	verb	noun	modifier	
					manner adverb
Differentiated	4	verb	noun	adjective	manner adverb
Rigid	4/5	verb	noun	adjective	(manner adverb)
	5	verb	noun	adjective	
	5/6	verb	noun	(adjective)	
	6	verb	noun		
	6/7	verb	(noun)		
	7	verb			

Figure 4: Parts-of-speech systems, including intermediate ones

An important point to be made is that the classification in Figure 4 is based on the properties of lexeme classes, not of word classes. Flexible lexemes, when put to use in a specific function, may receive inflections that are specific of that function. Thus, in the Turkish example (2) the lexeme *güzel* 'beauty', used as the head of a referential phrase, receives the possessive marker *-im* '1.POSS.' This possibility is lacking when the same lexeme is used as a modifier. The word *güzelim* 'my beauty' can thus be said to be a nominal word, but it is based on a lexeme that can be used flexibly in three different functions, each allowing different inflectional possibilities.

For further details on and argumentation for the approach to parts-of-speech systems outlined in this section see Hengeveld, Rijkhoff and Siewierska (2004).[2]

3 Four sets of predictions

The approach to PoS-systems outlined in section 2 leads to a number of predictions concerning the PoS-system of a language and other aspects of the grammar of that language. These predictions may be grouped together under four headings.

Identifiability
The more specialized a lexical class is, i.e. the more it is tied to one functional slot, the less it is necessary to mark this slot and the phrase it forms part of syntactically or morphologically, i.e. there is a trade-off between lexical structure on the one hand and syntactic and morphological structure on the other. An example of a prediction that follows from this observation is that rigid languages may be expected to display more freedom of word order than flexible languages.

Integrity
The formal integrity of a lexeme, i.e. its formal independence of morphological material specific to a certain function, increases its applicability in various functions. An example of a prediction that follows from this observation is that flexible lexemes may be expected not to show morphologically conditioned stem alternation.

Unity
The phonological, morphological and semantic unity of a lexical class increases its applicability in various syntactic slots. An example of a prediction that follows from this observation is that intrinsic gender and conjugation classes may be expected not to occur in flexible languages.

Pervasiveness
Flexibility and rigidity of lexical stems may be expected to correlate with functionality and rigidity of other morphological and syntactic units within the grammar and with functions not covered by the PoS-hierarchy. An example of a prediction that follows from this observation is that case-marked noun phrases or adpositional phrases may be expected to be used predicatively more readily in flexible languages than in rigid languages.

The following sections review the results obtained in earlier studies grouped together under these four headings.

[2] Hengeveld and van Lier (2008, 2010) propose a slightly different approach in which the predication-reference and head-modifier parameters interact in a two-dimensional grid. This model then predicts a number of further systems. These are not taken into account in the current paper.

4 Identifiability

4.1 Introduction

In languages with a differentiated or rigid PoS-system, classes of lexemes are tied to a specific functional slot. This fact facilitates the processing of the phrases that are headed by these lexemes. For instance, if a hearer comes across a noun, he is certain to have come across a referential phrase. In a flexible language, on the other hand, lexemes do not support processing in the same way. For instance, if in a flexible language a hearer encounters a lexeme that can be used as a modifier, the nature of the lexeme itself does not help to decide whether he has hit upon a modifier of a referential phrase or of a predicate phrase. One might expect then that in a flexible language other strategies have to be invoked to ensure successful communication. The alternative strategies available for the disambiguation of functions of flexible lexemes are constituent order and segmental marking. I will consider these separately at the clausal and phrasal levels in the following sections.

4.2 Clause

In languages that do not have a distinct class of verbs, i.e. type 1 and 1/2 in Figure 4, lexical information is insufficient to arrive at the identification of the predicate phrase and the referential phrases within a sentence, given that there are no separate lexical classes the members of which are used to fill the head slots of predicate phrases and referential phrases. Since the number of referential phrases in argument function in a sentence may vary, it is particularly the position of the main predicate that may help to disambiguate between the two types of phrases. Hengeveld, Rijkhoff and Siewierska (2004) therefore predict that in these languages the main predicate should occupy a uniquely identifiable position under all circumstances. Since only an initial and a final position in the sentence are uniquely identifiable, they predict languages of types 1 and 1/2 not to have predicate medial basic word order, unless the problem of identifying the constituents of the clause is solved by segmental means. This prediction is confirmed. The following examples from Samoan (Mosel and Hovdhaugen 1992: 52, 56) illustrate the phenomenon:

(11) *`Ua o tamaiti i Apia*
PERF go children LD Apia
'The children have gone to Apia.'

(12) *`O le maile sa fasi e le teine*
PRES ART dog PAST hit ERG ART girl
'The dog was hit by the girl.'

Samoan, a flexible language of type 1, has a predicate-initial basic word order. Deviation from this order is possible in the case of topicalization, as illustrated in (12), but in that case there is an explicit presentative marker such that the initial constituent can be interpreted correctly as not being the predicate. The same goes for the other languages of type 1 and type 1/2: they have predicate-initial or predicate final constituent order, and if they allow deviations from this order, this is marked explicitly through segmental means.

In the sample used by Hengeveld, Rijkhoff and Siewierska (2004) this also holds for languages of type 2 and type 2/3. The explanation for this is that languages of these types allow all kinds of non-verbal constituents to be used predicatively (see Hengeveld 1992b).

This again leads to further potential ambiguity as regards the interpretation of a constituent as a predicate phrase or a referential phrase, that can be solved by the same means as those listed above: rigid order and/or segmental marking. This is illustrated by the following examples from Turkish (Lewis 1967):

(13) *Yol* *uzun*
road long
'The road is long.'

(14) *uzun* *yol*
long road
'the long road'

The fixed constituent order patterns in Turkish, with the predicate in final position, helps identify (13) unequivocally as a clause, while (14) is interpreted as a phrase.

Further corroboration for the idea that languages with a flexible PoS-system have a more rigid syntax and morphology comes from the expression of semantic functions in flexible languages. Naeff (1998) studies the way in which languages express the semantic functions Recipient, Beneficiary, Instrument, Direction and Location in relation to their PoS-system. One of the options languages have is to use zero-marking for a specific relation, i.e. to use no marking at all. Naeff shows that languages of types 1 through 2/3 never use this option. Whether through head- or depentent marking, they will always use some strategy that signals the relationship explicitly. For example, the type 1/2 language Mundari marks these by postpositions when expressed by an independent referential phrase and in some cases within the predicative word when pronominal, while the type 1 language Samoan uses prepositions. Only in languages from type 3 onwards is zero-marking allowed.

4.3 Phrase

In all languages with some degree of flexibility, i.e. types 1 through 3/4 in Figure 4, there is potential ambiguity as regards the identification of heads and modifiers within and across predicate phrases and referential phrases. For instance, if a language has a class of flexible non-verbs and a speaker uses these to fill the head and modifier slot of a referential phrase, lexical information is insufficient to decide which one is the head and which one the modifier; and if a language has a class of flexible modifiers rather than separate classes of adjectives and manner adverbs, freedom of constituent order creates a situation in which an addressee does not know whether to interpret a lexeme as the modifier of e.g. a preceding noun or of a following verb.

On the basis of this observation, Hengeveld, Rijkhoff and Siewierska (2004) predict that in languages of type 1 through 3/4 (i) the order of head and modifier at the phrasal level is fixed within phrases, unless the problem of identifying head and modifier is solved by segmental means, so as to avoid ambiguity within phrases; and (ii) the order of head and modifier is consistent (i.e. modifiers of predicate phrases and referential phrases either both follow or both precede their head), unless the problem of identifying head and modifier us is solved by segmental means, so as to avoid ambiguity across phrases.

The prediction is borne out by the data: languages of types 1 through 3/4 have a fixed order of head and modifier or mark a deviation from this pattern segmentally[3], while languages of other types may or may not show such restrictions, and actually often don't. A case of a language not respecting the word order restriction but repairing this morphologically is Warao. Consider the following examples (Vaquero 1965: 50; Romero-Figeroa 1997: 71):

(15) *noboto* *sanuka*
child small
'small child'

(16) *Ma-ha* *eku* *ine* *yakera* *tane* *uba-te*
1.SG-POSS inside I beauty MANNER sleep-NPAST
'I sleep very well in my hammock.'

In Warao, a type 2 language, modifiers within referential phrases follow the head (15), while modifiers within predicate phrases precede their heads (16). The potential ambiguity arising from this is solved by the optional addition of the postposition *tane* 'manner', thus resolving the problem of functional ambiguity raised by its ordering patterns. It is characteristic of flexible languages that there is a need to do so.

4.4 Summary of correlations

The various properties of flexible languages that follow from the fact that constituents cannot be identified sufficiently on the basis of information that is intrinsic to the lexemes that are being used are summarized in Table 1. The top row lists the PoS-systems in order of increasing rigidity, the blank boxes in between numbers representing the intermediate types.

	1		2		3		4		5		6		7
Predicate initial or -final position (4.2)	Y				Y/N								
Overt marking of semantic functions (4.2)	Y				Y/N								
Fixed order of head and modifier (4.3)	Y						Y/N						

Table 1: Identifiability and PoS-system

5 Integrity

In languages with flexible lexemes, flexibility would be severely hampered if the shape of a lexeme would be sensitive to specific functional environments. For instance, one would not expect a contentive lexeme in a language of type 1 to exhibit suppletive forms for the plural when used as the head of a referential phrase: such a condition for suppletion would be useless in other environments, for instance when that same contentive is used as the modifier of a predicate phrase. Functional independence may be expected to be reflected in

[3] There is one potential counterexample to this claim. In Ngiti, a predicate-medial language, manner modifiers are placed in clause-initial or -final position, thus often not occurring contiguous to the verb. Here identifiability thus seems to be enhanced by extraposition.

formal independence, since the formal integrity of a lexeme increases its applicability in various functional slots.[4]

On the basis of these considerations Hengeveld (2007) hypothesizes that flexible lexemes may be expected not to show morphologically conditioned stem alternation, such as morphophonological variation, irregular stem formation, or suppletion. As an illustration, consider the following examples from Kisi (Tucker Childs 1995: 223, 243):

(17) a. *hûŋ*
come.HORT

b. *hûŋ* *lé*
come.HORT NEG

(18) a. *baa*
hang.HORT

b. *bee*
hang.HORT.NEG

In Kisi 'roughly 15% of all verbs exhibit ablaut' (Tucker Childs 1995: 241), often used to express the negative. The regular negation is illustrated in (17), while (18) illustrates the irregular negation. In the latter case a single word form expresses both lexical and grammatical content, as a result of which the stem cannot be identified separately. This is a morphological phenomenon that one would not expect in a language in which the lexemes involved are flexible, as the stem alternation is irrelevant in other functional environments.

The prediction outlined above amounts to saying that flexible stems will exhibit agglutinative or isolating morphology, and never fusional morphology. Since the degrees of flexibility vary from one flexible system to another, the exact predictions vary according to type of PoS-system:

- In languages of type 1 morphologically conditioned stem alternation will not occur with lexemes that may be used as heads of predicate phrases.
- In languages of types 1-2 morphologically conditioned stem alternation will not occur with lexemes that may be used as heads of referential phrases;
- In languages of type 1-3 morphologically conditioned stem alternation will not occur with lexemes that may be used as modifiers within referential phrases.
- (In languages of type 1-3 morphologically conditioned stem alternation will not occur with lexemes that may be used as modifiers within predicate phrases.)

The last prediction is given between brackets, as it cannot be tested, since only very few languages admit the expression of grammatical categories on manner expressions.

As shown in Hengeveld (2007), the languages of his sample confirm all three testable predictions, that is, no flexible stem in any of the languages studied exhibits fusional morphology. Flexible stems only participate in the morphological processes of agglutination and isolation. An interesting consequence of this conclusion is that it is not languages that should be classified in terms of their morphological type, but stem classes within languages.

The properties of flexible languages that follow from the fact that flexible stems need to be formally independent can be summarized as in Table 2.[5]

---

[4] See Plank (1998, 1999) for an insightful discussion of this correlation.

[5] A further result of the study reported on here is that in languages in which stem alternation does occur, the degree to which it is used can be systematically described using the PoS-hierarchy given in (10). For instance, if a language does not exhibit stem alternations for lexemes used as heads of predicate phrases, it will not exhibit stem alternation for any other class of lexemes; if it exhibits stem alternation for lexemes used as modifiers of referential phrases, it will also exhibit stem alternation for lexemes used as heads of referential phrases and as heads of predicate phrases; etc.

	1		2		3		4		5		6		7
Fusion – Head Predicate Phrase	N	Y/N											
Fusion – Head Referential Phrase	N			Y/N									
Fusion – Modifier Referential Phrase	N					Y/N							

Table 2: Integrity and PoS-system

Hengeveld (2007) furthermore shows that in languages that allow stem alternation, its presence or absence across functions can be predicted using the parts-of-speech hierarchy given in (10). If a language allows stem alternation with lexemes used in a function more to the right in the hierarchy, it will also allow stem alternation in functions more to the left in the hierarchy, and conversely. Verbs are thus the most likely candidates for stem alternation, followed by nouns, adjectives, and, trivially, manner adverbs.

6 Unity

6.1 Introduction

Flexible lexemes would lose much of their functional elasticity if the lexeme class they belong to is divided into subclasses, either phonological, morphological or semantic. The prediction would therefore be that differentiation within lexeme classes is absent to the extent that these classes are flexible. The following sections look at this issue from a morphological and a semantic perspective respectively.

6.2 Morphological subclasses

The morphological unity of a lexical class, i.e. the absence of intrinsic morphological subclasses (as opposed to semantic and phonological subclasses, see Corbett 1991) triggering specific morphological processes, increases its applicability in various functional slots. Taking this perspective, Hengeveld and Valstar (2010) hypothesize that this type of differentiation within lexeme classes will be absent in flexible languages. More specifically, they hypothesize that:[6]

- In languages without a true class of verbs (1-1/2), the lexical elements that are used as the head of a predicate phrase do not display conjugation classes;
- In languages without a true class of nouns (1-2/3), the lexical elements that are used as the head of a referential phrase do not display declination classes.

The languages of their sample fully confirm these predictions, even more so than expected, in the sense that for both hypotheses the generalization extends to one further PoS-type: languages of type 2 do not display conjugation classes, and languages of type 3 do not display declination classes.

[6] Note that these hypotheses are logically unrelated to the question whether there is stem alternation in a language or not. While stem alternation is often manifested in processes restricted to certain subclasses of lexemes, there may be stem alternation without lexical differentiation (as in the case of generally applicable morphophonological rules that affect the form of the stem in a fusional language), and lexical differentiation without stem alternation (as in the case of e.g. different suffixes for different subclasses of nouns in an agglutinating language).

6.3 Semantic subclasses

The semantic unity of a lexical class, i.e. the absence of semantic subclasses limiting the distribution of a lexeme increases its applicability in various functional slots too. The less internally differentiated a lexeme class is, the higher its degree of elasticity. This can be seen from examples such as the following ones from Mundari (Osada 1992: 89; Hoffmann 1903: 8, 100), discussed in Hengeveld and Rijkhoff (2005):

(19) *Dub-aka-n-a-e?*
sit-ASP-INTR-PRED-3.SG.SBJ
'He is still sitting.'

(20) *Hon dub-aka-d-i-a-e?*
child sit-ASP-TR-3.SG.OBJ-PRED-3.SG.SBJ
'He has caused a child to sit down.'

Mundari, a language of type 1/2, lexemes are not intrinsically intransitive or transitive, they are simply unspecified for transitivity. In specific uses their transitivity is therefore encoded separately: by means of the intransitive marker *-n* in (19) and the transitive marker *–d* in (20). The existence of both members of the pair shows that these markers do not detransitivize or transitivize, they simply indicate in what syntactic configuration the lexeme is being used.

Rijkhoff (2003) shows that in fact languages without a distinct class of verbs (i.e. types 1 and 1/2) never exhibit differentiation according to transitivity within their flexible lexeme classes, while languages with a distinct class of verbs always exhibit this differentiation. This confirms the suggestion above that the functional elasticity of a lexeme class does not combine very well with internal semantic differentiation within that class.

Another example of this is provided in Rijkhoff (2000, see also Rijkhoff 2004), in which it is shown that in flexible languages without a differentiated class of nouns (i.e. types 1 through 2/3), the flexible lexemes are always transnumeral, i.e. not intrinsically specified for singular (or plural) number. A morphosyntactic reflex of this is that, when containing a numeral, the referential phrase is not simultaneously marked for number.[7] Consider the following examples from Turkish (Lewis 1967: 26):

(21) a. *ada*
island
'island, islands'

b. *ada-lar*
island-COLL
'islands'

c. *on iki ada*
ten two island
'twelve islands'

The unmarked non-verb *ada* 'island' (21a) can be interpreted as either singular or plural, when followed by the suffix *–lar* 'collective' only the plural reading is available, and when preceded by a numeral (21c) the collective suffix is absent.[8]

[7] Rijkhoff (2004: 100-121) therefore argues that 'number' markers in these languages actually express nominal aspect rather than number.

[8] A further reflex of transnumerality, as noted in Rijkhoff (1993) is that verbal agreement with plural (or rather: collective) subjects is often singular. This is under certain conditions also true for Turkish (see Lewis 1967: 246).

It the light of the foregoing discussion is not surprising to find that flexible lexeme classes that may be used as the head of a referential phrase lack intrinsic coding of number. This way their flexibility of being used as modifiers of referential phrases and predicate phrases is not hampered in any way by intrinsic semantic features potentially incompatible with those functions.

6.4 Correlations

The properties of flexible languages that follow from the fact that flexible stems need to be undifferentiated both morphologically and semantically can be summarized as in Table 3.

	1	2	3	4	5	6	7
Conjugation classes	N	Y/N					
Intrinsic transitivity	N	Y					
Declination classes	N		Y/N				
Intrinsic number	N		Y/N				

Table 3: Unity and PoS-system

7 Pervasiveness

7.1 Introduction

So far the discussion has concentrated on lexical stems, both basic and derived, and their use in four different defining functions. In this section I would like to expand the perspective in two different directions. The first concerns the question to what extent flexibility applies at different levels, more specifically the root, stem, and word levels (Haig 2006, Don, Hengeveld and van Lier 2008, Lehmann 2008, van Lier 2009). The second concerns the question to what extent flexibility applies for functions other than the four defining ones that constitute the PoS-hierarchy. This section is subdivided accordingly.

7.2 Levels of analysis

7.2.1 Introduction

The PoS-hierarchy and the resulting classification presented above are based on a consideration of the behaviour of stems, both basic and derived. A number of recent papers, Haig (2006) being the first of these, have argued for considering the issue of flexibility at successive levels of morphosyntactic analysis. Haig (2006) more specifically argues for a principle of increasing categorization, i.e. decreasing flexibility, and Lehmann (2008) pursues this same issue independently. The levels of analysis that may be considered include at least the root, the basic stem, the derived stem, and the morphosyntactic word. The distinction between root and basic stem is especially relevant in those languages that manifest dependent roots, i.e. lexical roots that cannot occur but in combination with other roots or derivational affixes, and will not be considered here for lack of data from a substantial sample of languages.[9] The step from basic stems to derived stems is relevant to the relation between lexicon and syntax that is

[9] But see for instance Alfieri (fc.) for a study of this issue in Arabic, where the distinction between roots and stems is particularly prominent.

the central concern of this article, and will be addressed in 7.2.2. Once inserted into a morphosyntactic position basic and derived stems acquire the status of morphosyntactic words that are part of phrases and clauses. I will limit the discussion of these to flexibility in the use of phrases as predicates in 7.2.3 and that of the use of dependent clauses in various functional slots in 7.2.4.

7.2.2 Derivation

Smit (2007) studies the issue of lexical derivation in languages with different PoS-systems and shows that flexible languages often have derivational processes that create derived lexemes that are one step less flexible than the basic stems of the language. An example of this is given in (22a-b):

(22) a. *Hij spreek-t zacht / zacht-jes*
he speak.PRES.3.SG soft / soft-ADVR
'He speaks softly.'

b. *een zacht / *zachtjes oppervlak*
a soft / soft-ADVR surface
'a soft surface'

Dutch, a type 3/4 language, has basic lexemes, such as *zacht* 'soft', of the flexible modifer class, that can be used as modifiers of predicate (22a) and referential (22b) phrases, but there are derived lexemes, such as *zachtjes* 'soft-ADVR', that can only be used as modifiers in predicate phrases and thus belong to the class of manner adverbs. In a similar way, Turkish, a type 2/3 language, has non-verbs as basic lexemes, but there are derived lexemes that belong to the class of modifiers. And finally, Mundari, a type 1/2 language, has contentives as basic lexemes, but there are derived lexemes that are non-verbs, i.e. are flexible except for the fact that they cannot be used predicatively.

It is furthermore important to note that in all these cases the source of the derivational process is the next higher category in the parts-of-speech hierarchy.

7.2.3 Predication

There are large differences between languages as regards the kinds of units that can be used predicatively. Hengeveld (1992b) shows that these differences can be described in terms of a predicability hierarchy. The category of units that is least easily predicable on that hierarchy is that of possessive phrases. Consider the following examples from Imbabura Quechua (Cole 1982) and Yagaria (Renck 1975):

(23) *Chay wasi ñuka-paj-mi*
DEM house 1-POSS-FOC
'That house is mine.'
'That house is of me.'

(24) *M-igopa gagae' igopa-vie*
DEM-land your land-INT
'Is this land your land?'

In Quechua possessive phrases may be used as a non-verbal predicate, as shown in (23). In Yagaria this is not the case: possessive phrases can only be used attributively and cannot be predicated directly. This problem is circumvented in (24) by turning a noun phrase with a possessive modifier into a non-verbal predicate.

Hengeveld (1992b) shows that languages with flexibility in their PoS-system, i.e. types 1 through 3/4, consistently allow possessive phrases to act as a non-verbal predicate, while all languages with some degree of rigidity consistently never allow this. Differentiated languages sometimes do and sometimes do not allow the predicative use of possessive phrases.

A further issue that is of interest in relation to non-verbal predication is the way in which languages treat their non-verbal predicates from a formal perspective. Compare the following examples from Lango (Noonan 1981: 45) and !Xũ (Köhler 1981: 599):

(25) *Mán ‘gwôk*
DEM 3.SG.dog.HAB
‘This is a dog.’

(26) *Tʃ’ù zè: kè.fiè:*
DEM new hut
‘This is a new hut.’

In both examples a non-verbal predicate is used without the intervention of a copula, yet there is an important difference between the two. In Lango non-verbal predicates are inflected in precisely the same way as verbal predicates. In !Xũ, while verbal predicates are inflected regularly, the predicate in non-verbal predications is simply juxtaposed with its argument. In Hengeveld (1992b) the first strategy is called Ø1, the second Ø2.

One would expect the Ø1 strategy, that makes no difference between types of predicates, to be more typical of flexible PoS-systems, and this is indeed the case. Hengeveld (1992b) found that the Ø1 strategy is never used in languages with PoS-systems of types 5/6 through 7, whereas both Ø1 and Ø2 may be found in less rigid systems.

7.2.4 Subordination

A remarkable fact about certain flexible languages is that the flexibility they exhibit in their PoS-system shows up in their subordination system as well. Consider the following examples from Turkish (Göksel and Kerslake 2005: 423, 424), a type 2/3 language:

(27) *Orhan-ın bir şey yap-ma-y-acağ-ı belli-y-di-Ø*
Orhan-GEN INDEF thing do-NEG-V-NMLZ.IRR-3.SG.POSS obvious-V-PAST-3.SG
‘It was obvious that Orhan wouldn’t do/wasn’t going to do anything.’

(28) *Fatma-‘nın yarın gör-eceğ-i film*
Fatma-GEN tomorrow see-NMLZ.IRR-3.SG.POSS film
‘the film that Fatma is going to/will be seeing tomorrow’

In these examples the same type[10] of subordinate construction is used both as a complement clause (27) and as a relative clause (28), that is, a single construction type is

[10] The differences in form are phonologically conditioned.

used both as the (complex) head of a referential phrase and as a modifier within a referential phrase.

Van Lier (2009; see also van Lier 2006, Hengeveld and van Lier 2008) shows that languages with PoS-systems 1 through 2/3 always have at least some subordinate construction that is flexible as well, though possibly with a lower degree of flexibility than the flexible PoS in those languages. This ties in neatly with the more general observation made in 7.2.1 that flexibility decreases as complexity increases.

7.3 Further functions

A number of studies have dedicated themselves to the question whether for functions other than the ones covered by the PoS-hierarchy generalizations can be drawn as to their potential flexibility and/or rigidity.

Salazar-García (2008) shows that degree modifiers in Romance languages can be classified in different groups as regards their flexibility. An important observation is that degree modifiers of verbs are consistently more flexible than degree modifiers of adjectives and adverbs, i.e. modifiers of modifiers are more rigid than modifiers of predicates. Though his sample does not allow for crosslinguistic generalizations, the initial results are promising and worth investigating on a larger scale. Since the Romance languages show a high degree of specialization in their parts-of-speech systems, this research shows that further differentiation between systems is possible if further functions are taken into account.

Fleur (1999) studies the existence and nature of lexical locative and temporal modifiers in languages with different PoS-systems, and finds that in languages with PoS-systems 1 through 2/3 the lexical elements that can be used in these functions are flexible in nature. Thus, Turkish has many lexemes that can be used as the head of a referential phrase and as a locative modifier such a *ileri*, 'front, forward'.

As regards relator lexemes, Naeff (1998), in her study of the expression of semantic functions in relation to PoS-systems, notes that the use of serial verb constructions as a means of introducing participants is typical[11] of languages with PoS system 5 and higher, pointing at a relation between rigidity and serialization, and at the centrality of verbs in rigid languages.

7.4 Correlations

The properties of flexible languages that follow from the fact that the flexibility/rigidity of basic PoS-systems may be extended to other areas of the grammar and the lexicon can be summarized as in Table 4.

[11] In her sample there is one counterexample, Hmong Njua, which interestingly also shows up as an initial counterexample in Rijkhoff (2000).

	1	2	3	4	5	6	7
Non-verb derivation from flexible source	Y	N					
Modifier derivation from flexible source	N	Y	N				
Mann.adv. derivation from flexible source	N		Y	N			
Predicability of possessive phrase	Y				N		
Ø1-strategy	Y/N					N	
Flexible dependent clauses	Y		Y/N				
Flexible Locative and temporal modifiers	Y		Y/N				
SVCs introducing participants	N				Y		

Table 4: Pervasiveness and PoS-system

8 Conclusions

The cumulative results of the research reported on above may now be listed as in Table 5. In this table numbers following parameters refer to the relevant sections in this paper. Vertical bold lines show the relevant cut-off points between contiguous PoS-systems. These cut-off points just by themselves provide evidence for the relevance of the distinctions between all PoS-systems, including all intermediate ones, up to the distinction between type 5 and 5/6. Horizontal bold lines indicate which group of properties holds for a certain basic PoS system plus the next intermediate one. The groups of properties thus identified are cumulative, i.e. the properties under the highest horizontal bold line (i.e. all properties) hold for languages of types 1-1/2, the properties under the next horizontal bold line hold for languages of types 2-2/3, etcetera.

	1	2	3	4	5	6	7
Fusion – Head Predicate Phrase (5)	N	Y/N					
Conjugation classes (6.2)	N	Y/N					
Intrinsic transitivity (6.3)	N	Y/N					
Non-verb der. from flexible source (7.2.2)	Y	N					
Fusion – Head Referential Phrase (5)	N	Y/N					
Modifier der. from flexible source (7.2.2)	N	Y	N				
Predicate initial or –final position (4.2)	Y	Y/N					
Overt marking of semantic functions (4.2)	Y	Y/N					
Declination classes (6.2)	N	Y/N					
Intrinsic number (6.3)	N	Y/N					
Flexible dependent clauses (7.2.4)	Y	Y/N					
Flex.locative and temporal modifiers (7.3)	Y	Y/N					
Fusion – Modifier Referential Phrase (5)	N	Y/N					
Fixed order of head and modifier (4.3)	Y	Y/N					
Mann.adv.der. from flexible source (7.2.2)	N	Y	N				
Predicability of possessive phrase (7.2.3)	Y	N					
SVCs introducing participants (7.3)	N	Y/N					
Ø1-strategy (7.2.3)	Y/N	N					
	1	2	3	4	5	6	7

Table 5: Grammatical and lexical properties of languages with different PoS-systems

What can be learned from this inventory is that:
- the more flexible a language is in its use of lexemes, the more rigid it is in its syntax and morphology;
- the more flexible a language is in its use of lexemes, the more resistent it is to fusional morphology;
- the more flexible a language is in its use of lexemes, the more it lacks intrinsic lexical features, be they morphological or semantic in nature;
- the more flexible a language is in its use of lexemes, the more it is flexible in its use of phrases and clauses.

A typical flexible language thus:
- is predicate-final or initial;
- is agglutinative or isolating;
- has lexemes that are not specified for transitivity, number, conjugation class or declination class;
- is not only flexible in its use of lexemes but also in its use of phrases, clauses, and various types of adjuncts.

The PoS-system of a language can thus indeed be seen as a basic typological determinant. But it is also clear from the above that this is the more so the higher the degree of flexibility of the PoS-system involved.

Appendix 1. Samples used in the studies reported on

Source / *Language*	Fleur (1999)	Hengeveld (1992b)	Hengeveld (2007)	Hengeveld, Rijkhoff and Siewierska (2004)	Hengeveld and Valstar (2010)	Lier (2009)	Naeff (1998)	Rijkhoff (2004)	Rijkhoff (2003)	Smit (2007)
!Xũ		x								
Abkhaz	x	x	x	x	x	x	x	x	x	
Abun						x				
Alamblak			x	x	x	x		x	x	
Arabic, Egyptian		x								
Arapesh, Mountain			x	x	x	x		x	x	
Babungo		x	x	x	x	x	x	x	x	x
Bambara		x	x	x	x	x		x	x	
Basque	x	x	x	x	x	x	x	x	x	
Berbice Dutch	x		x	x	x	x	x	x	x	
Burmese								x	x	
Burushaski	x	x	x	x	x	x	x	x	x	
Cayuga								x	x	
Chinese, Mandarin		x	x	x	x	x	x	x	x	
Chukchi	x	x						x	x	
Dhaasanac						x				
Dutch		x						x	x	x
English										x
Galela								x	x	
Garo			x	x	x	x				
Georgian				x		x			x	
Gooniyandi						x				
Guaraní	x	x	x	x	x	x		x	x	
Gude	x		x	x	x			x	x	
Hausa		x								
Hdi						x				
Hittite			x	x	x			x	x	
Hixkaryana		x	x	x	x	x	x	x	x	
Hungarian	x	x	x	x	x	x		x	x	
Hurrian				x	x			x	x	
Ika								x	x	
Itelmen			x	x	x	x				
Jamaican Creole		x								
Japanese				x		x				
Kambera						x				
Kayardild	x		x	x	x	x	x	x	x	
Ket	x	x	x	x	x	x		x	x	x
Kharia						x				
Kisi			x	x	x	x		x	x	
Koasati			x	x	x	x		x	x	

Korean								x	x	
Krongo	x	x	x	x	x	x		x	x	
Lango	x	x	x	x	x	x	x	x	x	x
Lavukaleve						x				
Ma'di						x				
Mam		x								
Miao		x	x	x	x	x	x	x	x	
Mundari			x	x	x		x			x
Nahali		x								
Nama	x		x	x	x	x	x	x	x	
Nasioi		x	x	x	x			x	x	
Navaho	x	x	x	x	x		x			
Ngalakan		x	x	x	x			x	x	
Ngiti			x	x	x			x	x	x
Ngiyambaa		x								
Nivkh	x	x	x	x	x	x			x	
Nung	x		x	x	x	x		x	x	
Nunggubuyu			x	x	x	x		x	x	
Oromo			x	x	x		x	x	x	
Paiwan			x	x	x	x				
Pipil	x	x	x	x	x	x	x	x	x	
Polish	x		x	x	x	x				
Quechua, Imbabura	x	x	x	x	x	x	x	x	x	x
Sahu			x		x					
Samoan	x		x	x	x	x		x		x
Santali						x				
Sarcee								x	x	
Slave						x				
Sumerian		x	x	x	x			x	x	
Tagalog		x		x		x				
Tamil	x	x	x	x	x	x	x	x	x	
Thai		x				x				
Tidore				x						
Tsou								x	x	
Turkish	x	x	x	x	x	x	x		x	x
Tuscarora	x		x	x	x	x	x			
Vietnamese		x						x	x	
Wambon	x		x	x	x	x	x	x	x	
Warao			x	x	x	x				x
West Greenlandic	x	x	x	x	x	x		x	x	
Yagaria		x								
Yessan-Mayo		x								

References

Alfieri, L. (forthc.). 'The Arabic parts-of-speech system'.

Cole, P. (1982). *Imbabura Quechua* (Lingua Descriptive Studies 5). Amsterdam: North Holland.

Corbett, G. (1991). *Gender* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

Don, J., Hengeveld, K. and Lier, E. H. van (2008). 'Flexibility in grammar: Its diversity and its limits.' Paper presented at *Syntax of the World's Languages III*, Berlin, September 25-28, 2008.

Fleur, F. (1999). *Lexicale tijds- en plaatsuitdrukkingen: Achtentwintig talen onderzocht*. MA thesis, University of Amsterdam.

Göksel, A. and Celia, K. (2005). *Turkish: A comprehensive grammar.* London and New York: Routledge.

Haig, G. (2006). 'Word-class distinctions and morphological type: agglutinating and fusional languages reconsidered.' Manuscript Universität Kiel.

Hengeveld, K. (1992a). 'Parts of speech', in M. Fortescue, P. Harder and L. Kristoffersen (eds.), *Layered structure and reference in a functional perspective* (Pragmatics and Beyond New Series 23), 29-55. Amsterdam: Benjamins. Reprinted in M. Anstey and J. L. Mackenzie (eds.) (2005). *Crucial readings in Functional Grammar* (Functional Grammar Series 26). Berlin and New York: Mouton de Gruyter, 79-106.

--- (1992b). *Non-verbal predication: theory, typology, diachrony* (Functional Grammar Series 15). Berlin and New York: Mouton de Gruyter.

--- (2007). 'Parts-of-speech systems and morphological types.' *ACLC Working Papers* 2-1: 31-48.

Hengeveld, K., Rijkhoff, J. and Siewierska, A. (2004). 'Parts-of-speech systems and word order', *Journal of Linguistics* 40-3: 527-570.

Hengeveld, K. and Rijkhoff, J. (2005). 'Mundari as a flexible language', *Linguistic Typology* 9-3: 406-431.

Hengeveld, K. and Lier, E. H. van (2008). 'Parts of speech and dependent clauses in Functional Discourse Grammar', *Studies in Language* 32-3: 753–785.

--- (2010). 'An implicational map of parts-of-speech', *Linguistic* Discovery 8-1: 129-156.

Hengeveld, K. and Valstar, M. (2010). 'Parts-of-speech systems and lexical subclasses', *Linguistics in Amsterdam* 3-1: 1-20.

Hoffmann, J. (1903). *Mundari grammar*. Calcutta: The Secretariat Press.

Köhler, O. R. A. (1981). 'La langue !Xũ', in J. Perrot (ed.), *Les langues dans le monde ancien et moderne*, vol. 1. Paris: Centre National de la Recherche Scientifique, 557-615.

Lehmann, C. (2008). 'Roots, stems and word classes', *Studies in Language* 32-3: 546-567.

Lewis, G. L. (1967). *Turkish Grammar*. Oxford: Oxford University Press.

Lier, E. H. van (2006). 'Parts-of-speech systems and dependent clauses: a typological study', *Folia Linguistica* 40-3&4: 239-304.

--- (2009). *Parts-of speech and dependent clauses: A typological study* (PhD dissertation, University of Amsterdam; LOT dissertation series no. 221). Utrecht: LOT.

Mosel, U. and Hovdhaugen, E. (1992). *Samoan reference grammar* (Instituttet for sammenlignende kulturforskning, Series B: Skrifter, LXXXV). Oslo: Scandinavian University Press.

Naeff, R. (1998). *De expressie van semantische functies in typologisch perspectief: Een onderzoek naar de expressie van Recipient, Beneficiary, Instrument, Direction en Location in 22 talen.* MA thesis, University of Amsterdam.

Noonan, M. P. (1992). *Lango* (Mouton Grammar Library 7). Berlin and New York: Mouton de Gruyter.
Osada, T. (1992). *A reference grammar of Mundari*. Tokyo: Institute for the Languages and Cultures of Asia and Africa.
Plank, F. (1998). 'The co-variation of phonology with morphology and syntax: a hopeful history', *Linguistic Typology* 2-2: 195-230.
--- (1999). 'Split morphology: How agglutination and flexion mix', *Linguistic Typology* 3: 279-340.
Reh, M. (1985). *Die Krongo-Sprache (nìino mó-dì). Beschreibung, Texte, Wörterverzeichnis.* (Kölner Beiträge zur Afrikanistik 12.) Berlin: Reimer.
Renck, G.L. (1975). *A grammar of Yagaria* (Pacific Linguistics B40). Canberra: The Australian National University.
Romero-Figueroa, A. (1997). *A reference grammar of Warao* (Lincom Studies in Native American Linguistics 6). München: Lincom.
Rijkhoff, J. (1993). '"Number" disagreement', in A. Crochetière, J-C. Boulanger and C. Quellon (eds), *Proceedings of the XVth International Congress of Linguists*. Sainte-Foy, Québec: Presses de l' Université Laval, 274-276.
--- (2000). 'When can a language have adjectives? An implicational universal', in P. M. Vogel and B. Comrie (eds), *Approaches to the Typology of Word Classes* (Empirical Approaches to Language Typology 23). Berlin and New York: Mouton de Gruyter, 217-257.
--- (2003). 'When can a language have nouns and verbs?' *Acta Linguistica Hafniensia* 35, 7-38.
--- (2004; revised and extended paperback edn.). *The noun phrase*. Oxford: Oxford University Press.
Salazar García, V. (2008). 'Degree words, intensification, and word class distinctions in Romance languages', *Studies in Language* 32-3: 701-726.
Smit, N. (2007). 'Dynamic perspectives on lexical categorisation', *ACLC Working Papers* 2-1: 49-75.
Tucker Childs, G. (1995). *A grammar of Kisi* (Mouton Grammar Library 16). Berlin and New York: Mouton de Gruyter.
Vaquero, A. (1965). *Idioma Warao: Morfología, Sintaxis, Literatura.* Caracas: Editorial Sucre.

Derivation and categorization in flexible and differentiated languages

Jan Don and Eva van Lier

1 Introduction

Languages across the world seem to differ with respect to the way in which their lexical items are categorized. Recently, so-called *flexible* languages such as Mundari have reignited the discussion about the universality of the noun-verb distinction. One of the most prominent aspects of this discussion hinges on the semantic interpretation of words in purportedly flexible languages.

Two basic viewpoints regarding this issue have been expressed in the literature. On one view, if the same lexical form can be used in both nominal and verbal function, but the meanings of the lexemes in either function are not entirely predictable from that function, then the two meanings in fact correspond to two different lexical items, categorized as noun and verb, respectively (Evans and Osada 2005; Croft 2005). On the second view, lexical forms that may appear in nominal and verbal function have vague meanings that are compatible with both these functions. The actual usage of the lexeme as a verb or a noun profiles specific aspects of this vague meaning, and this may yield an interpretation that is idiosyncratic if compared to the meaning of another lexeme in the same function, or of the same lexeme in another function (Hengeveld and Rijkhoff 2005). Thus, defenders of the first view see semantic compositionality as a necessary condition for 'true' flexibility, while those that argue in favor of the second view believe that semantic irregularity is routinely – although not necessarily – expected in languages without distinct lexical classes of nouns and verbs.

The aim of this paper is to shed light on the debate about the (un)predictable nature of semantic shift in flexible languages. In particular, we aim to show that the two above-mentioned views are not mutually exclusive, since in purportedly flexible languages we encounter evidence for both of them. That is to say, we find cases of fully compositional interpretation of lexemes in verbal and nominal functions, as well as cases of more idiosyncratic semantic shifts. We will argue that the former, compositional type of interpretation involves syntactic derivation and is category-assigning. In contrast, unpredictable interpretation involves lexical derivation, and may or may not be category-assigning. Our central claim is that in flexible languages lexical (zero-)derivation of roots does not involve categorization; syntactic categorization occurs only at the phrase-structural level. In contrast, in the more familiar *differentiated* languages lexical derivation of a root has a categorizing effect, i.e. it determines the phrase-structural potential of the output form.

The paper is organized as follows. First, in section 2, we will further explain the two views on semantics in flexible languages; their assumptions and their implications. In relation to these, we will explicate our own assumptions – regarding storage and modification of lexical items and their meanings – of which we will make use throughout the paper. Subsequently, in sections 3, 4 and 5, we analyze a set of data from three supposedly flexible languages (Kharia, Tagalog, and Samoan), and compare these, in section 6, to data from a supposedly differentiated language (Dutch). On the basis of our findings, we argue in the discussion in section 7 that in flexible languages root-derivation and root-categorization are separated, while in differentiated languages they go together. Moreover, we will argue for the relevance of an additional factor involving complexity. In particular, we suggest that complex phrases are more likely to receive compositional semantic interpretations than simple lexemes, independent of whether they are syntactically categorized or not.

2 Categorization and interpretation

As mentioned in the introduction, two basic views on the semantic interpretation of flexible lexemes can be distinguished. The first view has been expressed most explicitly by Evans and Osada (2005; see also Croft 2005). These authors formulate the so-called *Compositionality Criterion*, stating that 'any semantic differences between the uses of a putative "fluid" [i.e. *flexible*, JD and EvL] lexeme in two syntactic positions (say argument and predicate) must be attributable to the function of that position' (Evans and Osada 2005: 367). If this criterion is not satisfied, i.e. if the semantic interpretation of a flexible lexeme in verbal versus nominal function is not entirely compositional, then each of the two meanings of a 'flexible' lexeme must be separately stored and labeled for the function to which they belong. This amounts to the same type of categorization that is found in differentiated languages and thus undermines the very notion of flexibility.

This view on semantics in flexible languages stands in contrast to Hengeveld and Rijkhoff 's position (2005; see also Rijkhoff 2008). They argue that idiosyncratic meaning shifts accompanying the verbal versus nominal usage of flexible lexemes may indeed be expected. According to them, flexible lexemes are semantically vague, and 'it is the use of a […] lexeme in a certain context (syntactic slot) that brings out certain parts of its meaning.' (Hengeveld and Rijkhoff 2005: 415). The exact constellation of highlighted meaning components that make up the interpretation of a lexeme in a particular function is not necessarily fully predictable. Note, however, that this does not exclude the possibility of semantically regular interpretations. Rather, Hengeveld and Rijkhoff's approach is compatible with both regular and more idiosyncratic semantic relations between the meanings of flexible lexemes used in different functions.

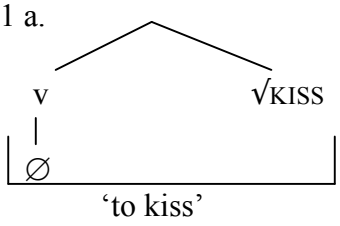
Thus, while defendants of both views hold that the interpretation of a flexible lexeme depends on the context in which it is used, they differ in whether or not this interpretation should always be fully predictable. These different viewpoints are crucially dependent on assumptions concerning the meaning of lexical roots. According to Evans and Osada, roots are stored with fixed meanings. In order for a language to be truly flexible, the semantic *coercion effect*, i.e. the semantics resulting from a particular syntactic function, should always be the same, independent of the meaning of the root used in that function (Evans and Osdada 2005: 270). This allows the meaning of the root and the semantics attributed by the syntactic function to add up to a compositional interpretation. In contrast, according to Hengeveld and Rijkhoff the flexibility of a flexible root precisely resides in the fact that its meaning is not fixed, but rather consists of various components that make it compatible with multiple functions. As said, the actual usage of a lexeme in a particular function profiles some aspects of this meaning, while usage in another function would yield a different type of profiling. Ultimately, however, the set of meaning components and the ways in which they may be foregrounded or backgrounded are lexeme-specific; they cannot be compositionally derived on the basis of the syntactic context.

We believe that the question whether the semantics of roots are vague or fixed is an empirical one, which remains as yet unanswered. In the remainder of this paper, however, we will assume that roots have inherent meanings. By this we intend that they have either a basic action-meaning or a basic object-meaning, while this does not imply that they are inherently categorized as either a verb or a noun. In particular, we make a principled distinction between being a noun or a verb as a matter of syntactic potential, and denoting an action or an object as a matter of conceptual semantics. This is in agreement with the cornerstone of Croft's universal-typological theory of parts of speech, according to which a prototypical noun is a lexeme combining object- semantics with referential function, and a prototypical verb is a lexeme combining action-semantics with predicative function (Croft

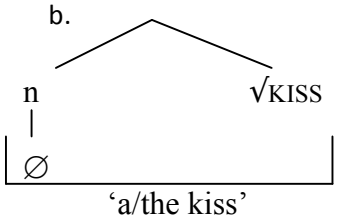
2001). Precisely the fact that meaning and function are separate dimensions in this theory allows for a mismatch between the two. In other words, syntactic categorization is in principle independent of denotational semantics.

Intimately related to the above is our assumption that, at a certain level of abstraction, not only flexible but also differentiated languages have un-categorized lexical roots (Marantz 1997, 2001). These roots are subjected to derivational processes, in which they combine with functional heads that – in differentiated languages – determine their syntactic category and their semantic interpretation as well. Thus, in English, we assume that an action-denoting root such as √KISS may undergo either a zero-marked derivational process that yields the meaning 'to kiss' and assigns the syntactic category of verb, or with a different zero-marked derivational process that yields the meaning 'a/the kiss' and assigns the syntactic category of noun. The structure in Figure 1a represents the verb 'kiss', while Figure 1b represents the noun 'kiss':

Fig.1 a.

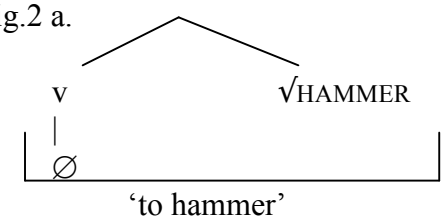


b.

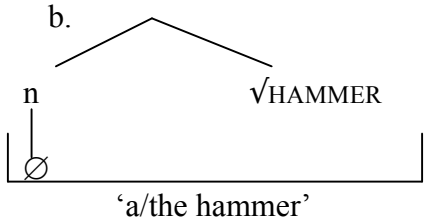


Alternatively, an object-denoting root such as √HAMMER may be combined either with a functional head – with zero spell-out – that assigns a nominal syntactic category along with the object-meaning 'a/the hammer', or with a zero functional head that gives it a verbal category and the semantic meaning 'to hammer'. This is represented in Figures 2a and 2b:

Fig.2 a.



b.



Crucially, this combination of a root with a first functional head involves a combination of lexical and syntactic (rather than purely syntactic) derivation, and as such it may yield unpredictable, idiosyncratic semantic interpretations. Thus, an action-denoting root combined with a nominal functional head may yield a semantic interpretation such as 'a particular instance of the action denoted by the root', as in the example with √KISS above. However, other, more or less idiosyncratic interpretations are possible with other roots; witness an interpretation such as 'the room where one carries out the action denoted by the root' as a nominal meaning derived from the root √STUDY, or an interpretation such as 'the result of the action denoted by the root' as a nominal meaning of the root √CUT. Similarly, when an object-denoting root is combined with a verbal functional head, this may yield the meaning of 'using the object denoted by the root as an instrument to perform a typical action with it', as in the example with √HAMMER above. Again, however, other roots may exhibit different types of semantic shifts.[1]

[1] See Farrell (2001) for an analysis of category-less roots in English based on Cognitive Grammar.

As we will argue in the remainder of this paper, flexible languages deviate from the procedure sketched above, in that they may subject an un-categorized root to a (potentially zero-marked) derivational process that yields an object-interpretation or an action-interpretation, without categorizing the output form as a noun or a verb in the syntactic sense. In other words, despite its noun-like or verb-like semantics, the resulting derived lexeme remains syntactically flexible: it can still be used – without further measures – in both a nominal and a verbal syntactic function, which involves a separate derivational step, namely the combination of the lexically derived lexeme with a particular syntactic functional head.

We will show that the first type of derivation in flexible languages, the type that has a semantic but not a syntactic effect on the base root, may involve unpredictable meanings, in the same way that root-level derivation in differentiated languages may be idiosyncratic, but without producing a categorized output form. On the other hand, we will show that the second type of derivational process in flexible languages, i.e. the type that provides flexible lexical material with a syntactic category by using it in a specific phrase-structural environment, always involves fully compositional semantic interpretations. Moreover, we will show that in cases of overt derivational processes (as opposed to zero-marked ones), semantic unpredictability co-occurs with phonological irregularity, while semantically compositional processes are also fully regular in their phonological form.

A final issue that needs to be covered before continuing with the analysis of actual language data is our use of zero-morphemes. As the above makes clear, we assume that derivational processes may involve zero-marking. However, in the case of lexical derivations, we assume zero-marking only when there is evidence, in the form of semantic shift, showing that a particular form has undergone some derivational process. In the case of syntactic derivation, we argue for zero-marked processes only when there is evidence that this zero form contrasts with other, overt forms in a particular paradigm of nominal or verbal marking, such as case or aspectual distinctions.[2]

3 Kharia

3.1 Flexible simple roots in Kharia

Kharia is an Austro-Asiatic language, belonging to the subfamily of South Munda languages. It is primarily spoken in the state of Jharkhand in India. According to Peterson (2006), Kharia is a flexible language. Specifically, Peterson (2005, 2006, this volume) claims that Kharia does not possess lexical classes of nouns and verbs, but instead has two *syntactically* defined categories, namely 'predicate' and 'complement' (Peterson 2006: 87). According to Peterson (2006: 60), 'lexemes in Kharia do not appear to be either inherently predicates or their complement but can generally appear in both functions, without any overt derivational morphology'.

The categories of predicate and complement are expressed through the combination of a 'content head' (Peterson 2006: 87) with a functional head. Content heads may consist of any single root, but also of multiple roots or complex phrases. All types of content heads can be combined with either a predicate or a complement functional head. The former is

[2] cf. Borer (2003), who presents a different view in which zero-affixes are non-existent. Borer argues on the basis of English data that if we assume that there are no zero-morphemes, the lack of 'eventive' interpretations of 'converted' nouns in English can be explained. These eventive interpretations only arise if a verb underlies these nouns; if zero morphemes do not exist, 'converted nouns' cannot be derived from verbs, hence their non-eventive interpretation.

expressed through enclitic markers of voice/tense and person[3], the latter through enclitic case marking (other than possessive/genitive).[4]

The semantic interpretation of a content head in an actual utterance requires a combination of the intrinsic meaning of the root(s) that it consists of, its syntactic function, and the morpho-syntactic distinctions belonging to that function (such as the voice markers in predicative function). According to Peterson (2006: 70), this combination yields '*entirely predictable*' semantic outcomes. In particular, if an object- (or individual-) denoting root is used as a predicate and marked for middle voice, it gets the meaning 'to become X', where X is the meaning of the root. When the same type of lexeme in the same function is marked for active voice it will be interpreted as 'to turn something into X'. Conversely, action-denoting roots that are used in referential function get the meaning '(the act of) X-ing', where X is again the meaning of the root.[5]

Consider now the examples in (1) and (2) below. Example (1) shows an individual-denoting root or content head, combined with a complement functional head in (1a) and with a predicate functional head in (1b). Example (2) shows an action-denoting root or content head, combined with a complement functional head in (2a) and a 'predicate' functional head in (2b). These combinations of root-meaning and function all yield the predicted, compositional semantic interpretations.

(1) a. ***lebu*** *ɖel=ki*
man come=MID.PST
'The/a man came.'

b. *bhagwan* ***lebu****=ki*
God man=MID.PST
'God became a man.' (= Jesus)

(2) a. *u* ***kayom*** *onɖor=kon raʈa=ya? ayo=ɖom,*
this talk hear=SEQ Rata=GEN mother=3.POSS
daʈhi=ya? saw=ʈay=ɖom gam=te:…
Darhi=GEN spouse=woman=3.POSS say=ACT.PRS
'Hearing this talking, Rata's mother, Darhi's wife says: …'

b. *ni lebu=ki khoʈi=ki=te* ***kayom=ta=ki***
story man=PL village.section=PL=OBL talk=MID.PRS=PL
'The people tell [this] story in the villages.' (Peterson 2006: 60, 68)

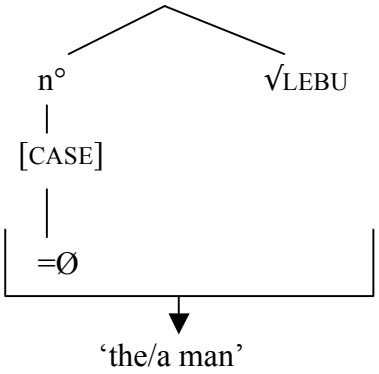
[3] Peterson (2006, this volume) distinguishes between the functional head of the predicate, spelled out by tense/voice, and the projection of the subject, spelled out by the person marker. Below, we will represent only the predicative head (spelled out by tense/voice markers).

[4] Subjects, non-definite countable objects, and non-countable objects take zero case. Definite, countable objects take the oblique (accusative/dative) case-marker *=te* (Peterson 2006: 109). In our analysis, we follow Peterson in assuming that nominal functional heads are spelled out by case. Alternatively, one may assume that the n-head has zero spell-out and that case is added as an inflectional category. This is a theoretically motivated choice, which does not have any consequences for our analysis of flexibility in Kharia.

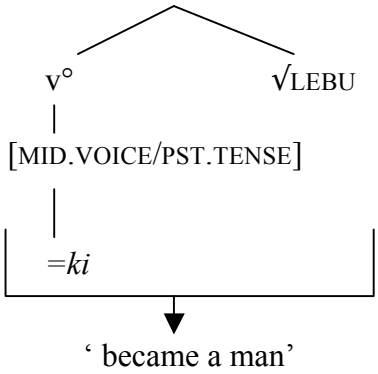
[5] Possibly, the different voice-markings correspond to different 'little v's' that result in the different semantic interpretations (cf. Folli and Harley (2007)).

Our analysis of the relevant parts of the data in (1a) and (1b) is given in Figures 3a and 3b respectively; note that we interpret Peterson's complement and predicate functional heads as nominal (n°) and verbal (v°) functional heads, respectively.

Fig. 3 a. N

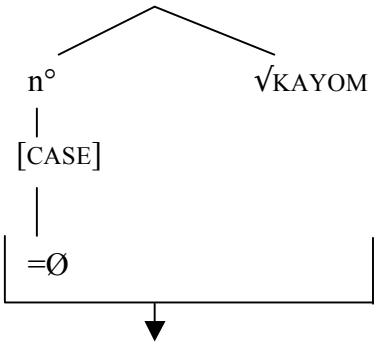


b. V

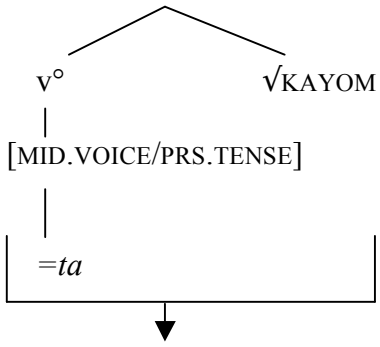


Similarly, the representations of the relevant data in (2a) and (2b) appear in Figures 4a and 4b:

Fig. 4a. N



b. V



3.2 Flexible derived roots in Kharia

Before combining with a functional head, Kharia roots may undergo an overtly marked lexical derivational process: so-called nasal-vowel-infixation. The vowel has the same quality as the vowel preceding the nasal of the infix; the nasal is of indeterminate quality: it is usually realized as /n/, but /m/ is also occasionally found. In line with this phonological irregularity, the semantic interpretation of the derived forms is unpredictable and idiosyncratic: they may denote objects, instruments or locations typically involved in the action denoted by their base root, but also specific instances or results of the action. (Peterson 2006: 79). Some examples of the process are provided in (3) (Peterson 2006:80):

(3)

	ROOT		DERIVED FORM	
a.	*bel*	'spread out'	*be-ne-l*	'bedding'
b.	*bui*	'keep, raise'	*bu-nu-i*	'pig'
c.	*jo?*	'sweep'	*jo-no-?*	'broom'
d.	*jin*	'touch'	*ji-ni-b*	'(a) touch'
e.	*rab*	'bury'	*ra-na-b*	'burial ground'

f.	*kuj*	'dance'	*ku-nu-j*	'dance'
g.	*mu?siŋ*	'rise'	*mu-nu-?siŋ*	'east' (i.e. where the sun rises')

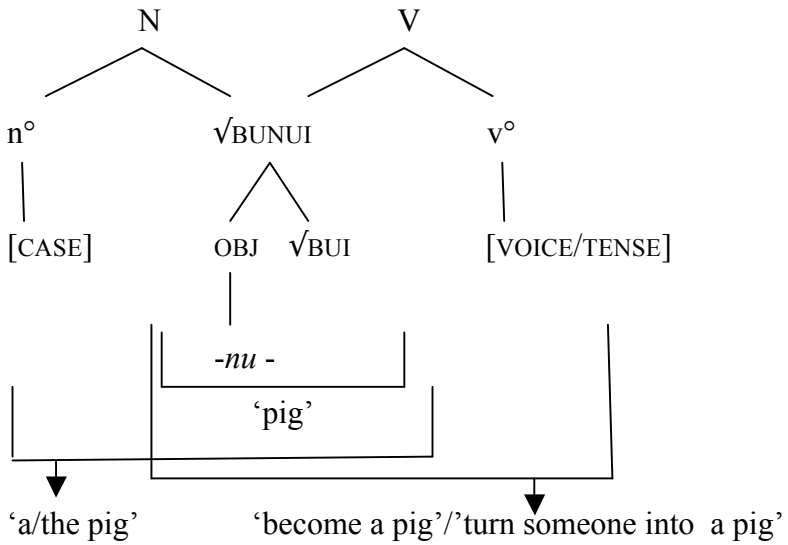
Both the semantic and the phonological properties of nasal-vowel-infixation are in accordance with the expectations for a root-level derivational process in a flexible language, which creates a different lexeme without imposing a syntactic category on it.

Note that we will refer to derived but un-categorized structures as derived *roots.* In this interpretation, the term *root* does not relate to the (lack of) internal complexity of a particular form, but rather to its syntactically un-categorized status. As we will see below, apart from derived roots, flexible languages also exhibit *root phrases*, i.e. combinations of multiple (derived) lexemes into complex but syntactically uncategorized structures.

The non-categorizing nature of nasal-vowel infixation in Kharia may seem counter-intuitive, since the semantics of the output forms appear to be more compatible with a nominal than with a verbal function (see (3) above). However, Peterson explicitly states that nasal-vowel-derived forms are flexible, i.e. that they can combine with both verbal and nominal heads, just like any other content heads (Peterson 2006: 81). Unfortunately, no examples are available to support the claimed flexibility of infixed forms. This lack of data is probably due to the fact that a suitable pragmatic context for predicative usage of these derivations will rarely be found.[6]

Despite this gap in the data, we propose, in accordance with Peterson's claim, an analysis of nasal-vowel-infixation in Kharia that is represented in Figure 5, taking the form in (3b) above as an example. Note that we represent the nasal-vowel-infix as OBJ for 'object', which is an informal means to indicate the semantic correlate of the derivational process, namely the formation of a lexeme that denotes 'an object that is in some way associated with the action denoted by the root'.

Fig.5



[6] Although there are no examples of these derived roots with 'nominal' semantics, Peterson (2005: 400) gives an example of a construction in which the underived object-denoting root *abul* 'table' is used as a predicate with active voice, yielding the compositional interpretation 'to turn someone into a table'. Another example is given in Peterson and Maas (2009: 215), where the root *boksel* 'sister in law' is used with middle voice, meaning: 'to become someone's sister-in-law' (cf. example (1b) above). In section 5.2 we discuss a very similar case in Samoan, in which we do find evidence for the verbal use of derived roots with - what may seem - 'nominal' semantics.

As can be seen in Figure 5, the root √BUI (‘keep, raise’) undergoes nasal-vowel-infixation, resulting in the derived root *bu-nu-i*, with the unpredictable semantic interpretation ‘pig’. This new lexeme is flexible, i.e. it can be combined with either a nominal or a verbal functional head. In the former case, the compositional semantic interpretation will be ‘a/the pig’; in the latter case it will either be ‘to become a pig’, or ‘to turn something/someone into a pig’, depending on whether the verbal head is spelled out by middle or active voice.

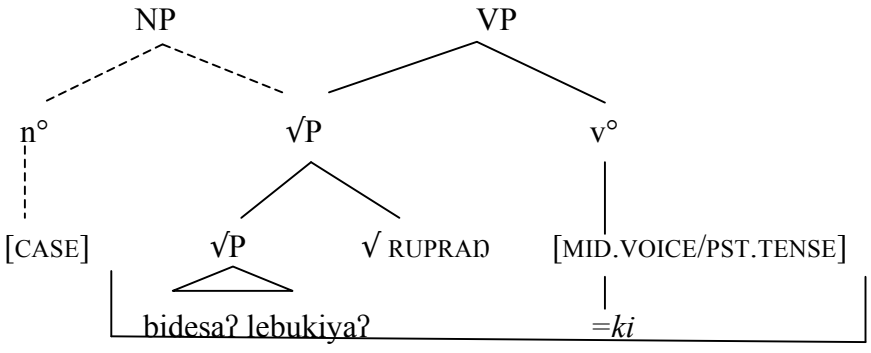
3.3 Flexible multiple-root phrases in Kharia

As mentioned above, the un-categorized units in Kharia that combine with a functional head need not be simple or derived roots, but may also be (and often are) complex combinations of multiple (derived) roots. An example of such a complex content head is given in (4); the whole phrase between square brackets (the translational equivalent of an English noun phrase) is combined with a verbal functional head, spelled out by middle voice:

(4) *bharat=ya lebu=ki* [*bides=a?* *lebu=ki=ya?* *rupraŋ*]*=ki=may.*
India=GEN person=PL abroad=GEN man=PL=GEN appearance=MID.PST=3PL
‘The Indians took on the appearance of people from abroad.’
(Peterson 2006: 85)

In Figure 6 we represent the structure of the relevant part of example (4). The root *rupraŋ* (‘appearance’) combines with the possessive root phrase (indicated simply as √P) *bidesa? lebukiya?* (‘of people from abroad’) into another complex root phrase. This complex phrase is then combined with a verbal functional head, spelled out by the voice/tense marker *=ki*. As expected, this syntactic derivation yields a semantically regular interpretation. Furthermore, the part with the dashed lines on the left-hand side of the figure is meant to indicate the flexibility of the root phrase (i.e. the part between square brackets in (4)): Rather than being combined with a verbal functional head, as in (4) above, the root phrase could instead have been combined with a nominal functional head, spelled out by case (see footnote 4).

Fig. 6



‘to take on the appeareance of foreigners’

3.4 Flexible 'nominalizations' in Kharia

We now turn to a second derivational process in Kharia, namely the formation of what Peterson calls 'freestanding forms' (Peterson 2006) or 'masdars' (Peterson and Maas 2009; Peterson this volume); we will employ the former term.

Phonologically, the formation of a freestanding form depends on the number of syllables of its base: freestanding forms of polysyllabic words (consisting of the root and possibly a causative marker) are zero-derived, whereas monosyllabic bases are reduplicated (Peterson 2006: 73). For instance, the freestanding form of a monosyllabic base like [*soŋ*] ('buy') is [*soŋsoŋ*], whereas the freestanding forms of polysyllabic words like [*obsoŋ*] ([CAUS-buy] 'sell') or [*kersoŋ*] ('marry') are phonologically identical to their base forms. Freestanding forms may take their own argument(s).[7] Semantically, the derivation of freestanding forms is regular: they are interpreted as denoting the event expressed by their base unit from a global perspective, i.e. without reference to its inherent temporal structure (Peterson 2006: 74).

Crucially, Peterson shows that freestanding form constructions are flexible, just like the output forms of the nasal-vowel-infixation discussed above. They can combine with either a verbal or a nominal functional head, as is illustrated in (5) and (6), respectively.

(5) *iɲ* [*ɖaʔ* ***biʔɖ-biʔɖ***] *=ki=ɲ*
1SG water pour.out-RDP=MID.PST=1SG
'I used to pour water out.' (i.e. that was my job).

(6) [*oʔ=yaʔ* ***bay-bay***] *um=iɲ* *baʔj=ta.*
house=GEN build-RDP neg=1SG like=MID.PRS
'I don't like (the act of) building houses.' (Peterson 2006: 73)

Freestanding forms thus represent a combination of properties that is unexpected from our current perspective. In particular, since freestanding forms are syntactically flexible, the process by which they are derived should be purely lexical, and as such susceptible to phonological and semantic irregularity. However, while the phonology of freestanding forms indeed varies depending on the form of the base root, the semantic interpretation of the output forms is apparently regular.

To account for this unexpected situation, we propose an additional, functional explanatory factor, which relates to the complexity of the base-form of a given derivational process. In particular, we argue that the semantic regularity of freestanding form constructions is due to the fact that the process applies to a construction denoting an event, which is typically expressed through a complex unit consisting of an action-denoting root plus root(s) denoting the participant(s) in this event. The functional motivation behind this explanation is that more complex units are more easily computationally processed than memorized as chunks.[8]

---

[7] Objects of freestanding forms take the same case marking as they do in independent clauses. What corresponds to the subject in an independent clause appears in the genitive case with a freestanding form, if it is expressed overtly. Occasionally, objects also appear in the genitive (Peterson 2006: 72; Peterson and Maas 2009: 225).

[8] Note that we expect that the effect of this functional factor will not be absolute. On the one hand, complex constructions may indeed be stored as chunks, with fixed idiomatic interpretations. On the other hand, simple

It should also be noted, however, that our understanding of the nature of non-categorizing derivational processes, as opposed to syntactically categorizing ones, does not force idiosyncratic semantic meanings; it merely allows for them. Therefore, the fact that we do not find evidence for irregularities in the interpretation of a particular type of output structure does not imply that such interpretations are ruled out in principle. Rather, if we do find non-compositional semantics, we interpret this as evidence in favor of our analysis. We realize, however, that this approach does not allow us to make empirically strong, falsifiable predictions on this point.

We may now consider the interpretation of freestanding forms in the larger context of an actual utterance. If a freestanding form is combined with a nominal functional head, as in example (6) above, this involves compositional interpretation. However, freestanding form constructions in verbal function can only take middle voice, and receive a habitual interpretation, as the translation of (5) above makes clear.[9] Thus, there is a semantic difference between a predicate that is formed on the basis of a freestanding form, and its equivalent, formed on the basis of a simple lexeme. Such a simple predicate structure is exemplified in (7):

(7) *iɲ* *ɖaʔ* *biʈh=oʔj*
1SG water pour.out:ACT.PST=1SG
'I poured water out.' (Peterson 2006: 73)

Peterson argues that the habitual interpretation of freestanding forms in verbal function is indeed predictable:

> 'I would argue that the habitual interpretation of such forms [i.e. freestanding forms, JD and EvL] in Kharia results from the fact that the reduplicated form is intimately connected to a depiction of the event without explicit reference to its internal temporal structure. If such a form is nevertheless marked as a finite predicate, what results is a habitual situation, not an activity or event in the usual sense [...].' (Peterson 2006: 74).

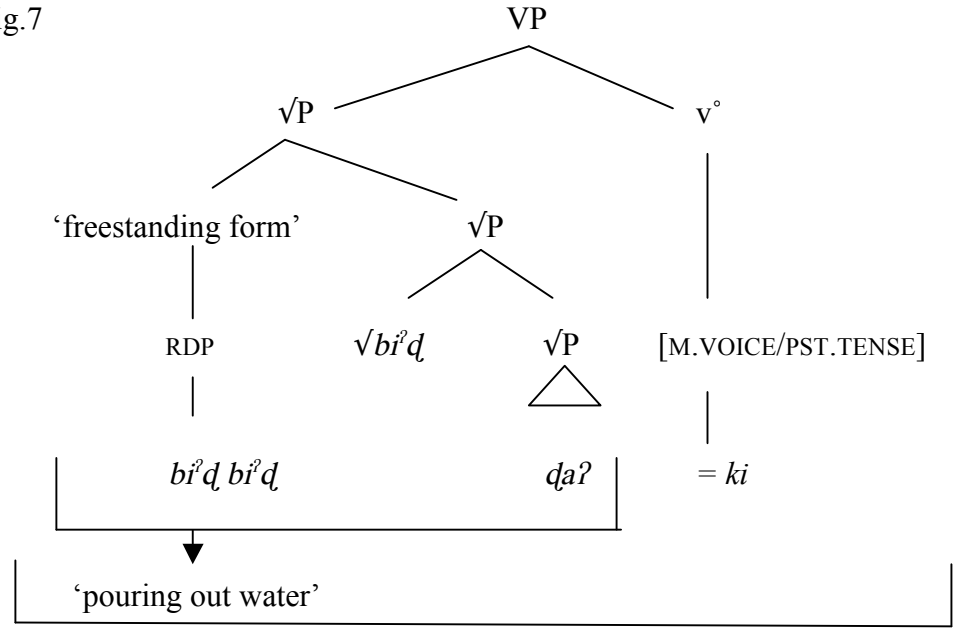
Thus, following Peterson, we contend that the syntactic derivation of freestanding form constructions yields compositional semantic interpretations. In Figure7 we give an analysis of the process, using the data from example (5) above:

---

flexible roots may, as a result of usage frequency in a particular function, become lexicalized and stored with an idiomatic interpretation linked to a syntactic category.

[9] In fact, polysyllabic, non-reduplicated freestanding forms used in verbal function with middle voice are interpreted as prolonged events or events in the distant past or distant/uncertain future. For this reason, Peterson (2006: 197) refers to this construction as the 'generic middle', rather than the 'habitual middle'. For examples and more discussion, see also Peterson and Maas 2009: 229-231).

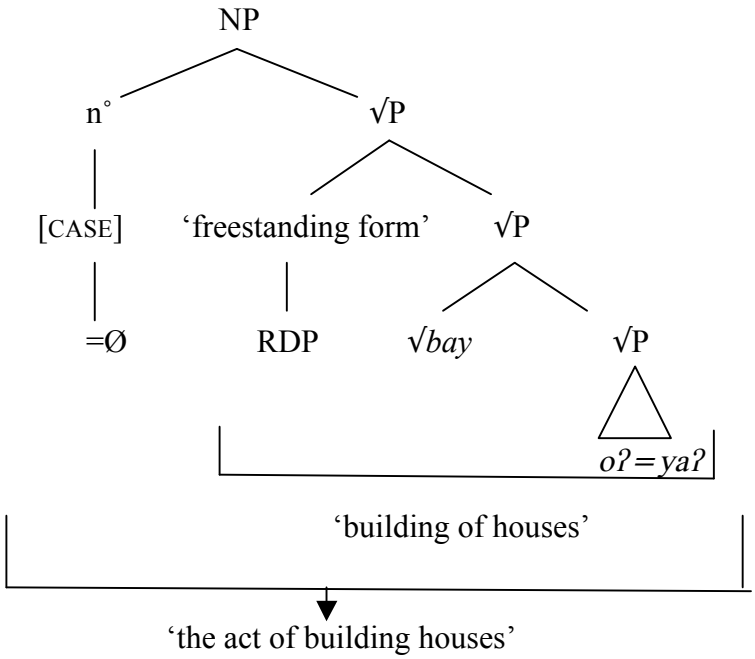
Fig.7



The root *bi?ɖ* (‘pour out’) is combined with its internal argument *ɖa?* ‘water’ into a root phrase. This root-phrase is then subjected to the freestanding form-deriving process. This involves phonological reduplication (RDP), as *bi?ɖ* is monosyllabic, and semantic interpretation as ‘the event denoted by the base from a global perspective’. Syntactically, the output of this derivation remains an uncategorized root phrase, which is subsequently combined with a verbal functional head, spelled out by the enclitic middle voice marker, which results in a predictable habitual interpretation.

Now consider Figure 8, which represents the nominal use of the freestanding form illustrated in example (6) above.

Fig.8



The root *bay* (‘build’) combines with its argument *o?=ya?* (‘of houses’), resulting in a root phrase. This root-phrase combines with the ‘freestanding form’ head – again spelled out by

reduplication, as *bay* is monosyllabic – and yielding another root phrase with the meaning 'building of houses'. This phrase, which at this point is still un-categorized, combines with a nominal functional head, spelled out by case, and is compositionally interpreted as 'the act of building houses'.

3.5 Summary

To summarize, category-less roots in Kharia can be derived and combined into complex phrases without losing their syntactic flexibility. We have shown that such processes typically involve both semantically and phonological unpredictable interpretations. As such, these processes stand in contrast to processes by which linguistic units – be they simple, derived, or otherwise complex – are combined with syntactically categorizing functional heads; the latter are semantically (and phonologically) fully compositional. A potential counterexample is presented by the formation of so-called 'freestanding forms', which are compositionally interpreted while being flexible at the same time. We have suggested that the semantic regularity of this process is due to the complex nature of its base form: freestanding forms are based on complex event-denoting structures and this makes them less amenable for idiosyncratic interpretation, since storage of complex structures is relatively costly compared to computation.

4 Tagalog

4.1 Lexical and syntactic zero-derivation in Tagalog

Tagalog is an Austronesian language spoken in the Philippines. Regarding the flexibility of the language, Himmelmann (2005: 361) remarks: 'Content words do not have to be sub-classified with regard to syntactic categories. They all have the same syntactic distribution, i.e. they all may occur as predicates [and] as (semantic) heads of noun phrases […]'. Thus, in Tagalog, as in Kharia, both action-denoting and object-denoting roots can combine both with a verbal and a nominal functional head.

According to Himmelmann (2007: 276), action-denoting roots in nominal function may receive one of the following semantic interpretations:

(i) the state which ensues from the successful performance of an action;
(ii) the result or the typical cognate object of the action;
(iii) the name of the action.

In what follows, we limit ourselves to the discussion of meanings (ii) and (iii).[10] Many roots, according to Himmelmann, can receive both these interpretations. This is illustrated in (8), where the root *lakad* 'walk' has meaning (ii) in (8a) and meaning (iii) in (8b).

(8)a. *Iyón ay ma-haba'-ng* ***lakad***
DIST PM STAT-length-LK walk
'That was/is a long walk

b. *Ma-husay ang* ***lakad*** *ng mákiná*
STAT-orderliness SPEC walk GEN machine
'The walking of the machine is good.' (Himmelmann 2007: 280)

[10] The stative meaning in (i) appears to be only relevant in adjectival function, which is beyond the scope of the present discussion.

We argue that the 'object/result'-interpretation in (ii) above and the 'action-name'-interpretation in (iii) above represent the outcomes of two structurally different types of zero-marked derivational processes: a purely lexical, un-categorizing process and a syntactic categorization, respectively. Our motivation for this analysis is as follows.

Firstly, the description of the 'object/result'-interpretation ('the result *or* the typical cognate object of the action') suggests that it potentially involves semantic irregularities, which are in fact reminiscent of those reported for nasal-vowel-infixation in Kharia. Moreover, we will see in sections 5.1 and 5.2 that Samoan exhibits the same type of root-level derivation with both zero and overt spell-outs, and with parallel semantic interpretations.[11]

Secondly and crucially, example (8a) above shows that in Tagalog a form such as *lakad* ('a walk'), with an 'object/result' meaning derived from an action-denoting root, can be used as a predicate despite its noun-like semantics. The syntactic categorization as a verb yields the expected, compositional semantic interpretation: 'to be a (long) walk'. Note that verbal syntactic categorization in Tagalog normally involves sentence-initial position. Whenever the predicate is *not* the first lexeme of the utterance, as in (8a) above, it is obligatorily marked by the functional element *ay*, glossed PM for *predicate marker*.

In contrast to the zero-derived forms with a 'result/object'-interpretation, we analyse derivations yielding the interpretation in (iii) above ('the name of the action') as resulting from a syntactic process that combines an action-denoting root with a nominal functional head, as is the case in (8b) above. We contend that Himmelmann's description of the semantic coercion effect 'the name of the action' is the equivalent of what Peterson describes for Kharia as 'the act of X-ing', i.e. the compositional semantics of an action-denoting root used in a nominal syntactic function.

Thus, we propose that there are two formally identical but structurally different, flexible forms of *lakad* in Tagalog. First, there is an *un*-derived category-less root $\sqrt{\text{LAKAD}_1}$, which denotes 'to walk' when used syntactically as a verb, and 'walking' when used syntactically as a noun. The latter usage is exemplified by (8b) above. The structural representations of this type of syntactic derivation, i.e. combining the root with a verbal or nominal head, is represented in Figure 9 below. Second, there is a zero-derived, still syntactically un-categorized form $\sqrt{\text{LAKAD}_2}$, which means 'a walk' (i.e. 'the result of the action of walking'), and which can be used as noun, but also as a verb with the compositional semantic interpretation 'to be a walk'. This latter usage is exemplified by (8a) above. In Figure 10 below, we represent the lexical zero-derivation of $\sqrt{\text{LAKAD}_2}$, indicating the operation with RES for 'result' as an approximation of the semantic effect of the process (in the same way that we used OBJ in the representation of the Kharia derivation in Fig 5 above). The upper part of the schema indicates the flexibility of $\sqrt{\text{LAKAD}_2}$, i.e. the fact that it can be combined with either a nominal or a verbal nominal head, and with compositional semantic results.

---

[11] Paolo Acquaviva (p.c.) points out that there is nothing idiosyncratic about the result-meaning 'a walk', derived from an action-denoting root 'walk'. We agree that this meaning shift appears quite regular, but at the same time we argue that it is crucially different from the compositional semantic interpretation '(the act of) walking' that accompanies syntactic categorization. Moreover, the meaning definition in (ii) above makes clear that the semantic shift in these cases is not always exactly the same: rather than the 'result' of an action, a zero-derived root may also have the meaning of an 'object' typically used in the action; this depends on the meaning of the base root. Nonetheless, Acquaviva's comment highlights the fact that semantic shifts resulting from lexical derivation are not necessarily 'quirky'; they may be quite expected, but this does not automatically make them compositional.

Fig.9

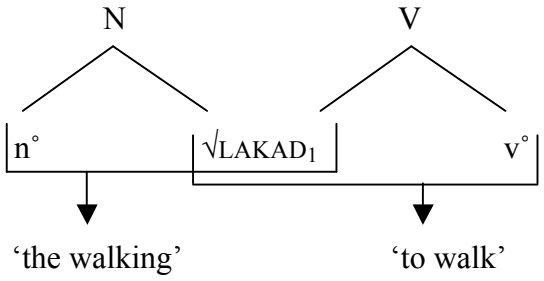
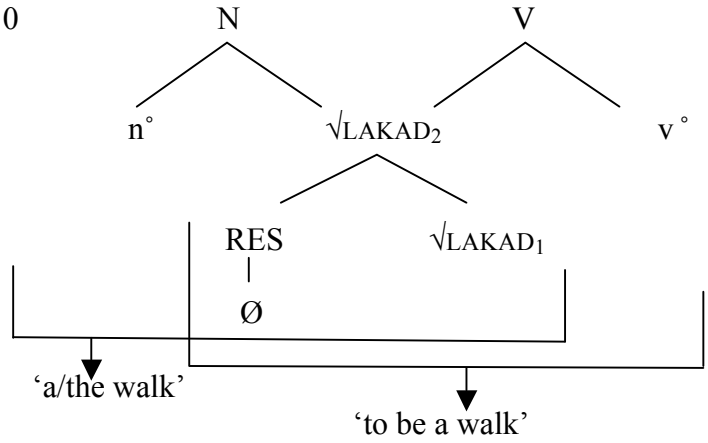


Fig.10



Note finally that the form *lakad* in (8a) above could alternatively be analyzed as a zero-derived noun, functioning as a non-verbal predicate without a copula. The argument supporting such an analysis would be that other (un-derived) roots in predicative function take voice marking in Tagalog. However, since the un-derived root $\sqrt{\text{LAKAD}_1}$ and the derived form $\sqrt{\text{LAKAD}_2}$ are homophonous, a voice-marked form of *lakad* would be interpreted as an action-denoting predicate, oriented towards a particular participant (i.e. 'someone walks'). In other words, the 'result'-semantics of the derived lexeme are incompatible with voice-marking in predicative function.

4.2 Voice-marked forms in Tagalog

As mentioned above, Tagalog un-derived roots (both action-denoting and object-denoting) are usually marked for voice when used in verbal function. According to Himmelmann (2007: 288), voice-marking is itself a derivational process that produces a new lexical item, denoting an 'actor- or undergoer-oriented action expression', which is still flexible in terms of its syntactic possibilities. Thus, Himmelmann analyses Tagalog voice marking as a process that does *not* involve syntactic categorization of the base root. As expected, the output forms of the voice-marking process are characterized by pervasive formal and semantic idiosyncrasies. In terms of phonology, derivation with voice markers may involve unpredictable deletion of root vowels, as illustrated in (9a), and, sporadically, insertion of /n/, as seen in (9b). There are also completely irregular forms, such as for example in (9c):

	ROOT	VOICE-MARKED FORM
(9)a.	*kain* 'consumption of food'	*kan-in* 'to eat something' (patient voice form)
b.	*tawa* 'laugh, laughter, laughing'	*tawa-n-an* 'to laugh at someone' (locative voice form)
c.	*kuha* 'getting, a helping'	*kun-in* 'to get something' (patient voice form)

Semantically, the interpretation of voice-marked forms is broadly based on the meaning of the root and the meaning of the voice marker, but again there are many idiosyncrasies. For instance, the root *anák* means 'child', whereas the derived form *mag-anak* does not mean 'to give birth', but 'to breed', and is most commonly used for animals (Himmelmann 2007: 288). In (10), we give some more examples of such unpredictable semantic interpretations of voice-marked forms derived from object-denoting roots:

	ROOT	VOICE-MARKED FORM
(10) a.	*bus* 'bus'	*mag-bus* 'ride a bus'
b.	*kamay* 'hand'	*mag-may-an* 'shake hands'
c.	*langgam* 'ant'	*langgam-in* 'be infested with ants'
d.	*lubid* 'rope'	*lubir-in* 'be made into rope' (Foley 1998)

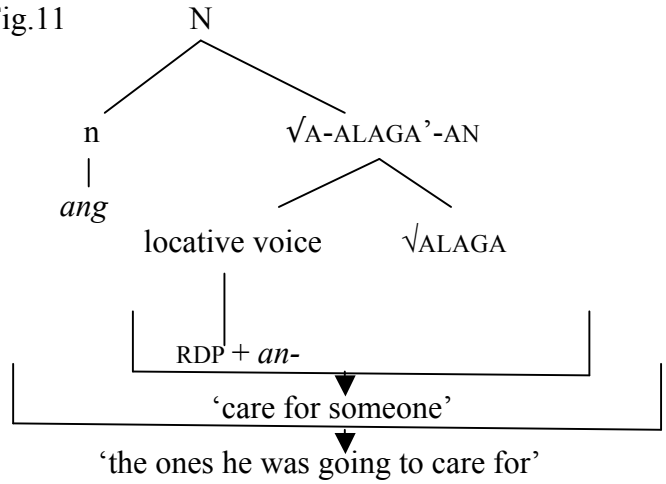
Notably, an alternative analysis of voice marking would be that it derives a verb from a category-less root, and that this verb is then zero-derived in case it is used in nominal function. Following Himmelmann, however, we argue that the flexibility analysis is to be preferred, because there is no difference between voice-marked forms and other (unmarked) roots in terms of their possibility to combine with both verbal and nominal functional heads. In other words, there is no evidence for an additional layer of categorization for voice-marked forms.[12] The flexibility of voice-marked forms is illustrated in (11), showing the nominal use of a form marked for locative voice:

(11) *i-u-uwi'* *niya* ***ang*** ***a-alaga'-an*** *nya*
CV-RDP-returned_home 3.SG.POSS SPEC RDP-cared_for-LV 3SG.POSS
'He would return the ones he was going to care for.' (Himmelmann 2007: 266)

The structure of the relevant form in (11) is represented in Figure 11 below. In the lower part of the structure, the root √ALAGA 'care for' is joined with the locative voice marker, spelled out by a combination of (partial) reduplication and the suffix *–an*. This produces the flexible derived root √A-ALAGA'-AN, meaning 'care for someone'. Subsequently, this voice-marked form is combined with a nominal functional head, spelled out by the function word *ang,* yielding the compositional semantic interpretation 'the one(s) he was going to care for'.[13]

[12] See Himmelmann (2007) for more discussion and additional arguments in favor of this analysis.
[13] Note that the object/individual semantics of the form in nominal function are dependent on its voice-marking. The use of an unmarked form *alaga* in nominal function would be interpreted as 'the act of caring'. See also the discussion further below.

Fig.11



As the above example illustrates, the syntactic categorization of voice-marked forms yields predictable semantic interpretations: An actor-voice form in nominal function denotes 'the agent of the action' (and 'the action oriented towards the agent' in verbal function); a patient-voice form in nominal function denotes 'the patient of the action' (and 'the action oriented towards the patient' in verbal function); and so on (Himmelmann 2007).

4.3 Gerunds in Tagalog

Finally, we consider gerund formation in Tagalog. This involves an overt derivational process, spelled out by a set of allomorphs of prefixes and reduplication. Gerunds have regular semantic interpretations: They 'refer to actions or states without orienting them towards one of the participants involved in the action or state via voice marking' (Himmelmann 2005: 372). However, gerund formation is claimed to have no effect on the category of its base structure, despite the action-nominal semantics of the resultant constructions. Thus, gerunds can still combine with either a verbal or a nominal functional head, as is illustrated in examples (12a) and (12b), respectively. In (12a) the verbal functional head is realized through sentence-initial position; in (12b) the nominal function head is spelled out by the locative marker *sa*, which belongs to a small, closed set of functional elements (including *ang*, as illustrated in (8b) and (11) above) that can be used to mark noun phrases in various sentential functions.

(12) a. [***pag-lu****-luto?* *ng pagkain*] *ang trabaho niyá*
GER-RDP-cook GEN food SPEC work 3SG.POSS
'His/her job is cooking food.'

b. *pag-bawal-an* *mo* *ang bata?-ngiyó*
SF-forbidden-LV 2SG.POSS SPEC child-LK DIST
sa [***pag****-la-laró?* *sa* *lansangan*]
LOC **GER**-RDP-play LOC street
'Forbid that child to play in the street.' (Himmelmann 2005: 372, 373)

The regular semantic interpretation of gerunds is unexpected, assuming that the process by which they are derived does not involve syntactic categorization. This situation may be explained by the same functional factor that we invoked to account for the regular semantics of Kharia freestanding forms. In particular, like Kharia freestanding forms,

Tagalog gerunds are formed from complex root phrases, rather than simple roots, and more complex operands are unlikely, because un-economical, candidates for idiosyncratic interpretation.

It may be noted that, as in the case of *lakad* in example (8a) above, the gerund construction in (12a) could alternatively be analyzed as a nominalization functioning as a predicate without a copula, since it does not bear voice-marking. However, as we have argued above, voice-marking in Tagalog is not category-determining. Verbal syntactic categorization is realized either through word order (sentence-initial position) or by means of the predicate marker *ay*. In addition, gerunds denote actions without orienting them towards one of the participants. As such, gerund constructions are semantically incompatible with voice-marking. For these reasons, the lack of voice-marking on the gerund form in (12a) does not imply that it is a syntactic nominalization.

In relation to the above, it is interesting to note that the allomorphy encountered in voice-marking systematically co-varies with the allomorphy of gerund formations, as is illustrated in (13) below. So a root that takes *-um-* as the active voice-marker takes *pag-* as a gerund prefix, whereas a root that takes *mag-* for active voice, takes *pag-* plus reduplication for gerund formation, etcetera. This co-variation gives support to the idea that voice-marking and gerund formation are mutually exclusive. It also suggests that Tagalog distinguishes different classes of roots, which select different (classes of) allomorphs, but which have no bearing on the syntactic distribution of their members.[14]

(13)

ACTIVE VOICE-MARKER	GERUND-MARKER
-um-	*pag-*:
mag-	*pag-* + reduplication:
maN-	*paN-* + reduplication:
ma-	*pag-*+*ka-*

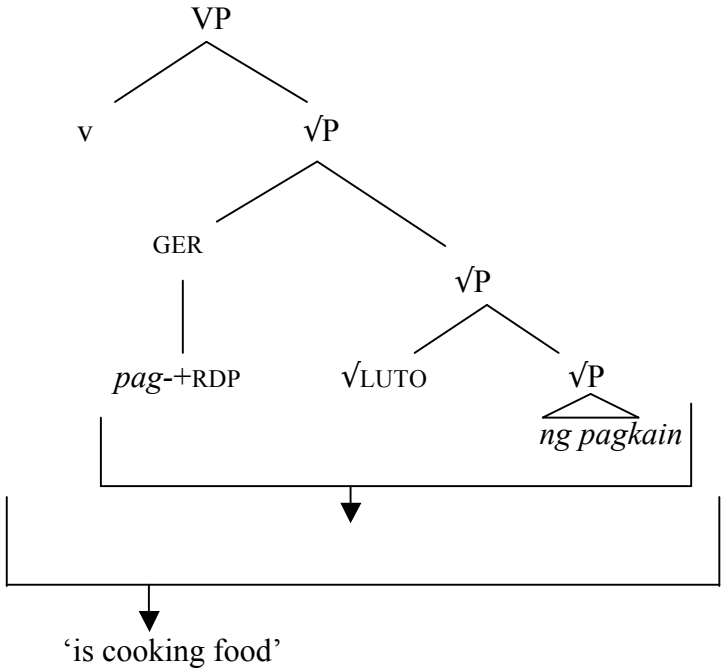
(Schachter and Otanes 1972: 160-161)

The structure of the bracketed gerund construction in (12a) above can thus be represented as in Figure 12 below. The action-denoting root √LUTO 'cook' combines with the possessive phrase *ng pagkain* ('of food') into a complex root phrase.[15] This root phrase is then subjected to the gerund-forming derivational process, indicated by GER for 'gerund', and spelled out by a combination of the prefix *pag-* and reduplication. This yields another root phrase meaning 'cooking of food', which is then combined with a verbal head (with no spell out, but syntactically marked though sentence-initial position). The syntactic categorization of the gerund form is compositionally interpreted as 'is (the) cooking (of) food'.

[14] Himmelmann (2007: 269) calls these 'morpho-lexical classes' as opposed to 'terminal syntactic categories'; only the latter type of classes are syntactic in nature, i.e. they are directly relevant for phrase-structure.

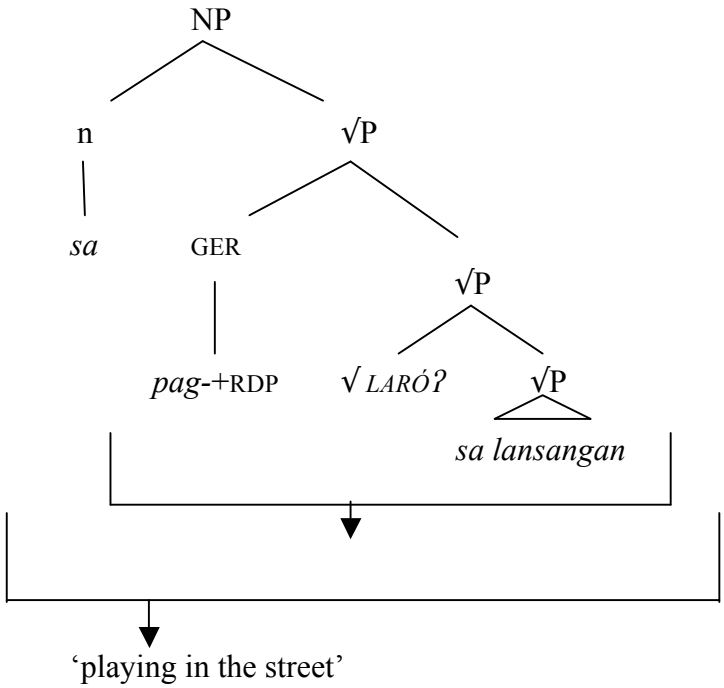
[15] The classification of the coding of the second argument in Tagalog gerunds is not entirely straightforward, because there is no difference between the marking of the second argument of an actor-voice predicate and the possessor in Tagalog: both are marked with *ng*. However, there is a second type of possessive construction in which the possessor is expressed as a *sa*-phrase. Since the first argument in a gerund construction can be both a *sa-* and a *ng*-phrase (just like possessors), while the second argument can only be a *ng*-phrase, Koptjevskaja-Tamm (1993: 119-120) argues in favour of a sentential (or *accusative*) analysis of the marking of the second argument.

Fig. 12



For the sake of completeness, the representation of (12b), with the gerund in nominal syntactic function, in provided in Figure 13: √*LARÓʔ* combines with the adjunct *sa lansangan* ('in the street)' into a root phrase, which is then subjected to gerund formation. The resulting root phrase, meaning 'playing in the street' is syntactically categorized by combining it with a nominal head spelled out by *sa*:

Fig. 13



4.4 Summary

In sum, in Tagalog, as in Kharia, category-less roots can be subjected to derivational processes, with overt or zero spell-out, and combined into complex phrases without losing

their syntactic flexibility. Such non-categorizing processes are likely to display semantic and phonological irregularities, in contrast to syntactic categorization, which yields compositional interpretations. However, non-categorizing derivational processes that apply to complex root phrases are semantically regular, presumably because the interpretations of complex constructions, categorized or not, tend to be computed rather than stored.

5 Samoan

5.1 Lexical and syntactic (zero-)derivation in Samoan

Samoan, like Tagalog, belongs to the Austronesian language family. According to Mosel and Hovdhaugen (1992: 77), 'in Samoan, the categorisation of words into nouns and verbs is not given a priori in the lexicon. It is only their actual occurrence in a particular syntactic environment which gives them the status of a verb or a noun'.

Thus, Mosel and Hovdhaugen take syntactic distribution as the primary criterion for flexibility. In particular, when a word is preceded by a TAM (Tense-Aspect-Mood) particle, it functions as a verb, and when it combines with one the following particles, it functions as a noun: *le* ('determiner', glossed as ART for 'article'), *e* (ergative marker; absolutive is zero-marked), *o* (possessive marker), or *i,* (marker of non-core arguments and locative-directional adjuncts).

Examples (14) and (15) provide evidence that the semantic interpretation of Samoan roots in verbal versus nominal syntactic function is completely regular: The action-denoting root √ALU 'go', used as a verb 'to go' in (14a), denotes 'the action of going' when used in nominal function in (14b). And the individual-denoting root √*UŌ* 'friend', used as a noun ('his friend') in (15a), is interpreted compositionally as 'be friends' when functioning as a verb with a plural subject, as shown in (15b).

(14) a. *E* ***alu*** *le* *pasi* *i* *Apia*
GENR **go** ART bus DIR Apia
'The bus **goes** to Apia.'

b. *le* ***alu*** *o* *le* *pasi* *i* *Apia*
ART **go** POSS ART bus DIR Apia
'the going of the bus to Apia'

(15) a. *E* ***uō*** *Tanielu* *ma* *Ionatana*
GENR **friend** Daniel and Jonathan
'Daniel and Jonathan **are friends**.'

b. *E* *alofa* *Taniel* *i* *l=a=na* ***uō***
GENR love Daniel DIR ART=POSS=3SG **friend**
'Daniel loves his **friend**.' (Mosel and Hovdhaugen 1992: 77)

Our analysis of the data in (14) and (15) is represented in Figures 14 and 15, respectively. In Figure 14 it is shown that √ALU can be combined with either a verbal functional head, spelled out by the general TAM-marker *e* in (14a), or a nominal functional head, spelled out

by the determiner *le.*[16] Both combinations have compositional semantic interpretations. Figure 15 represents the combination of √UŌ with either the verbal head *e* or the nominal head *lana* (a combination of a determiner and a possessive form), again with regular semantic outcomes.

Fig.14

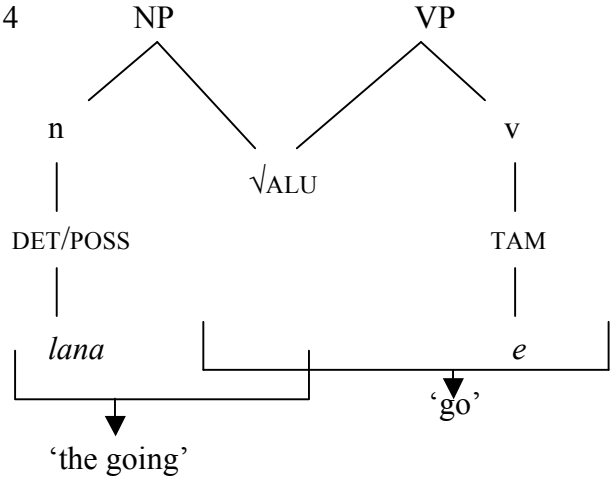
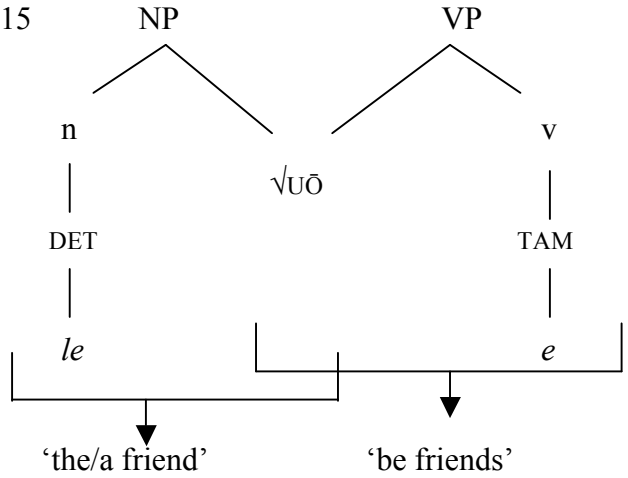


Fig.15



However, there is also evidence for idiosyncratic semantic shift in Samoan. Some examples of this are provided in (16)-(19) below (from Mosel and Hovdhaugen 1992: 82, 83).

(16) Actions or instruments with which these actions are performed:
a. *fana:* 'gun' or 'shoot'
b. *sasa*: 'rod, can, whip' or 'beat'
c. *lama*: 'torch' or 'fish by torch light'

[16] In line with our analysis of Kharia case marking, we assume that the determiner spells out the nominal function head. Another possibility would be to assume a zero nominal head, to which the determiner is added. As in the case of Kharia, the choice between these alternatives does not affect our argumentation.

(17) Actions or actors:
a. *gaoi* ‘steal’ or ‘thief’
b. *solo* ‘move forward’ or ‘procession’

(18) Actions or the results of these actions:
tusi: ‘write’ or ‘letter/book’

(19) Institutions or the state of belonging to this institution:
a. *ā’oga*: ‘school’ or ‘ attend school’
b. *eklaesia* ‘church’ or ‘be a church member’

Since the relation between the two meanings of the forms in (16)-(19) varies in unpredictable ways, we analyse these examples as instances of lexical zero-derivation. The output forms of this derivational process are new lexical items, which have different meanings but remain syntactically flexible. As such, the process is comparable to the zero-derivation of ‘object/result’-forms from action-denoting roots in Tagalog (cf. example (8a) above), and to overt nasal-vowel-infixation in Kharia (cf. (3) above).

Taking the form *tusi* from (18) above as an example, our claim is thus that there are two flexible, phonologically identical roots $\sqrt{\text{TUSI}_1}$ (un-derived) and $\sqrt{\text{TUSI}_2}$ (zero-derived from $\sqrt{\text{TUSI}_1}$). The semantic interpretation of $\sqrt{\text{TUSI}_2}$ is unpredictable: the fact that its meaning is ‘letter’ (rather than, for instance, ‘writer’, or ‘pen’) must be stored. Further, we claim that, despite its object-semantics, $\sqrt{\text{TUSI}_2}$ is still un-categorized. In other words: while $\sqrt{\text{TUSI}_2}$ denotes an object and while objects prototypically function as (syntactic) nouns, this does not mean that $\sqrt{\text{TUSI}_2}$ is a noun. Probably, it does mean that $\sqrt{\text{TUSI}_2}$ will not occur in verbal function very often (in the same way that e.g. Kharia *bunui* ‘pig’ will not). And in fact, we have no examples of verbal usages of forms like $\sqrt{\text{TUSI}_2}$. As we will show in the next subsection, however, we do have evidence of overt lexical derivations with noun-like, i.e. object-meanings, used in verbal function.

Assuming that both $\sqrt{\text{TUSI}_1}$ and $\sqrt{\text{TUSI}_2}$ are flexible, each form should be combinable with nominal and verbal categorial heads, and with compositional semantic interpretations: $\sqrt{\text{TUSI}_1}$ means ‘to write’ in verbal function and ‘writing’ in nominal function, while $\sqrt{\text{TUSI}_2}$ means ‘letter’ in nominal function and ‘be a letter’ in verbal function. We make this analysis explicit in Figure 16 for $\sqrt{\text{TUSI}_1}$ and in Figure 17 for $\sqrt{\text{TUSI}_2}$. Note that in Figure 17 we indicate the first derivational process by means of OBJ for ‘object’, with zero spell-out, in the same way as we used OBJ and RES in the representations in Figures 5 and 10 above. The spell-outs of the nominal and verbal functional heads have been omitted, as there are several possibilities.

Fig.16

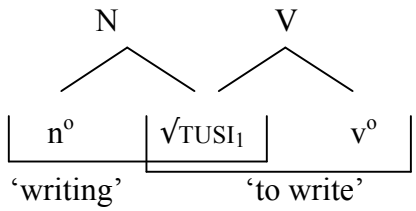
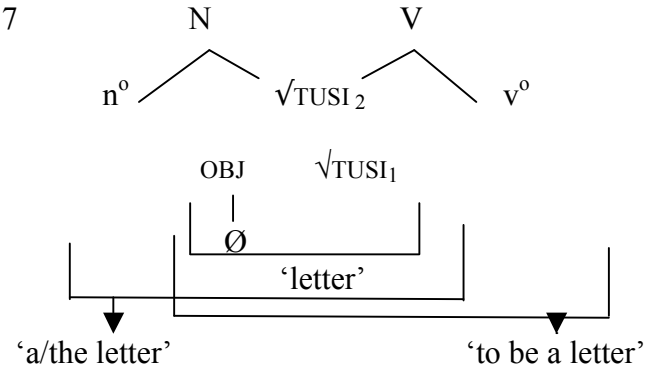


Fig.17



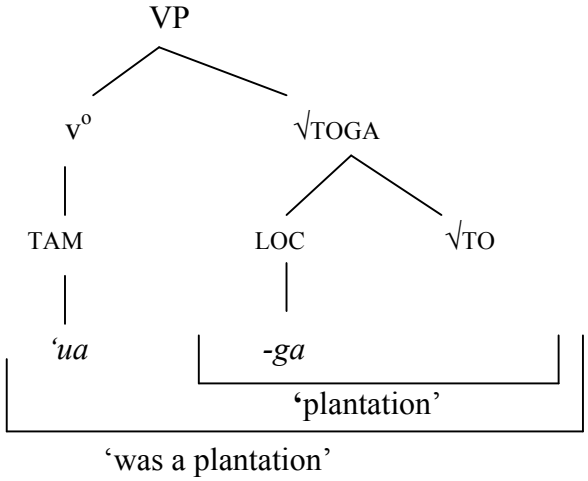
5.2 Overt lexical and syntactic derivation in Samoan

In the previous subsection, no data were available to illustrate the flexibility of lexical zero-derivations. Presumably, this is due to the pragmatic markedness of object-denoting lexemes in verbal function. In the present sub-section, however, we show that there are overtly derived counterparts of zero-derived forms, which also have 'nominal' semantics, but still combine freely with verbal functional heads, and with predictable semantic effects.

The relevant derivational process is spelled out by the suffix *-ga*, which is traditionally called a 'nominalization' suffix, obviously because of the noun-like semantics of the output forms. However, as Mosel (2004: 267) remarks, 'even words carrying the so-called nominalization suffix can occur in the head position of a verb complex'.This is illustrated in (20) and its representation in Figure 18, where the suffix *-ga* is combined with the root √TO 'plant', to yield the derived form √TOGĀ. This derived form has the unpredictable meaning of 'plantation', which may be roughly described as 'location associated with the meaning of the base', abbreviated as LOC in Figure 18. In example (20) √TOGĀ enters into a compound with *niu* 'coconut', but we have left this out of Figure 18 for the sake of clarity. Subsequently, √TOGĀ is combined with a verbal functional head, spelled out by the past tense marker *'ua*, and this yields a compositional semantic interpretation: 'was a (coconut) plantation'.

(20) *'Ua to-**gā-**niu 'ātoa le mea maupu'epu'e*
TAM plant-**NMLZ**-coconut whole ART place hill
'The whole hilly place was now a coconut plantation.' (Mosel 2004: 267)

Fig.18



Interestingly, the suffix *-ga* is used in Samoan to spell out derivational processes of different types. According to Mosel and Hovdhaugen (1992: 195), *-ga* derivations 'denote the object or the place involved in the event denoted by the verb [i.e. the *semantic* verb or action-denoting root, JD and EvL], or a specific type of event'. These types of interpretations are parallel to those of the lexical zero-derivations in examples (16)-(19) above. However, there are also cases in which *-ga* formations mean 'the act of X-ing', where X is the action denoted by the base root, parallel to the interpretation of *alu* '(the) going' in example (14b) above. Moreover, in terms of phonology, some *-ga* formations involve vowel-lengthening in the base form, while others do not.

We claim that the opposition between *ga*-formations with vowel-lengthening and those without vowel-lengthening corresponds to a difference between (i) lexical, semantically unpredictable and non-categorizing derivation, and (ii) semantically regular, syntactically categorizing derivation, i.e. nominalization. In (21) we give some examples of the semantic difference between non-categorizing *ga*-formations with vowel-lengthening and categorizing *ga*-formations without vowel-lengthening:

(21)

	ROOT	NON-CATEGORIZING	CATEGORIZING
a.	*amo* 'carry'	*āmo-ga* 'person(s) carrying loads	*amo-ga* 'carrying'
b.	*a'o* 'teach'	*ā'oga* 'school'	*a'o-ga* 'education'
c.	*tīpi* 'cut'	*tīpi-ga* 'surgical operation'	*tipi-ga* 'cutting'
d.	*pule* 'control'	*pulē-ga* 'unit of church administration'	*pule-ga* 'controlling'

(Mosel and Hovdhaugen 1992: 195)

We take the final item of (21d) as an example for the structural representations given in Figures 19 and 20 below. Figure 19 shows that the root √PULE 'control' can be combined with a nominal head, spelled out as *-ga*, which yields the predictable interpretation '(the act of) controlling'. The same root could alternatively be merged with a verbal head, spelled out by a TAM particle, with the regular meaning 'to control'. Figure 20 represents the non-categorizing *-ga* derivation, with vowel-lenthening, which yields the form √PULĒGA. This derived form has the idiosyncratic meaning 'unit of church administration'; the semantic shift produced by the derivation is informally indicated in the schema by the word 'institution'. Subsequently, √PULĒGA can combine with a nominal or a verbal head (the spell-outs of which are omitted for the sake of simplicity), and this should yield compositional semantic interpretations: 'the/a unit of church administration' in nominal function and 'be a unit of church administration' in verbal function. Note that the representations in Figures 19 and 20 are the overtly marked counterparts of those in Figures 16 and 17 above.

Fig.19

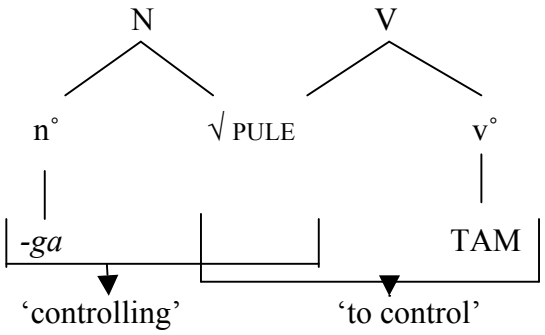
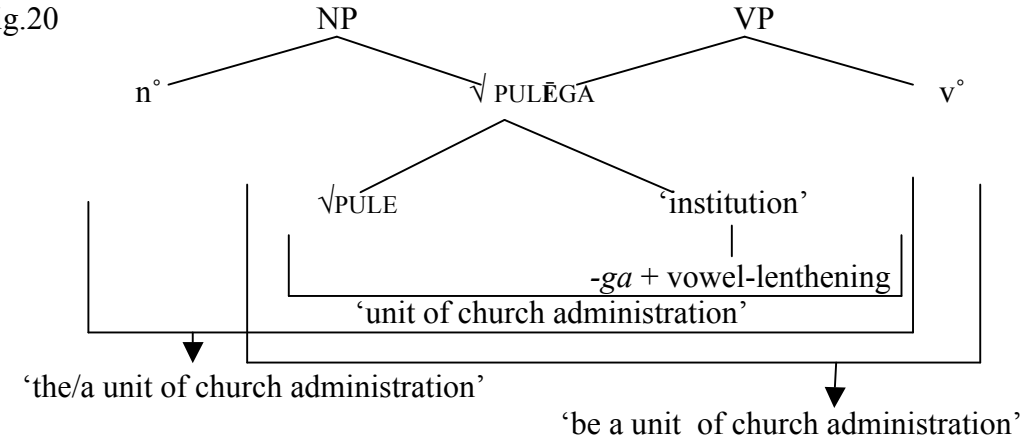


Fig.20



The double status of the *-ga* suffix in Samoan ties in with earlier studies on the differences between affixes that attach to words and affixes that attach to stems. Aronoff and Sridhar (1987) show that occasionally one and the same affix can be used at both levels, but triggering different semantic and phonological effects. More recently, Embick and Marantz (2008) claim that the distinction between stem-affixes and word-affixes may be reinterpreted structurally, such that stem affixes may give rise to idiosyncratic interpretations since they are in the same local domain as a root. Word-affixes, in contrast, are outside this local domain and hence no idiosyncratic interpretations may arise.

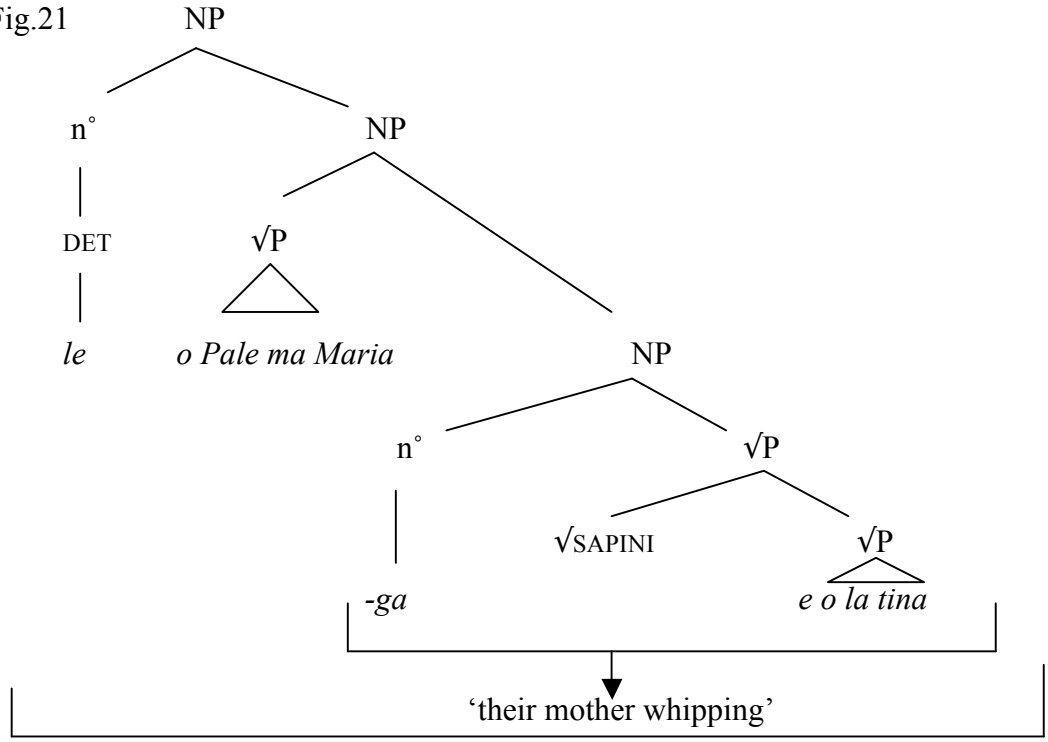
Notably, the Samoan suffix *-ga* is also productively used for syntactic nominalization of complex constructions. As we would expect, this process does not involve vowel lengthening and has regular semantics: 'syntactic nominalizations denote the action as a whole' (Mosel and Hovdhaugen 1992: 197). An example is provided in (22), the structural representation of which appears in Figure 21:

(22) *Ae na oo lava ina moumou malie atu le*
but PST reach EMPH CONJ disappear gentle DIR ART

pisa o le [*sapini=**ga** o Pale ma Maria*
noise POSS ART whip=**NMLZ** POSS Pale and Maria

e o la tina]
ERG POSS 3DUAL mother
'But finally the noise of the whipping of Pale and Maria by their mother faded away.' (Mosel and Hovdhaugen 1992: 57

Fig.21



'the whipping of Pale and Maria by their mother'

Importantly, these syntactic *-ga* nominalizations in Samoan are different from Tagalog gerunds and Kharia freestanding form constructions in that they are true nominalizations; they are restricted to usage in nominal function and cannot combine with a verbal head anymore.

On a final note, it is interesting to see that exactly the same two-level analysis that we propose for Samoan *-ga* derivation, can be applied to derivations with *-tanga* in Maori. In fact, the example in (23) shows the two types of derivation based on the same root √TANGI 'cry', in a single clause. The lexical, non-categorizing derivation has the idiosyncratic meaning 'funeral', while the syntactically categorizing (i.e. nominalizing) derivation is compositionally interpreted as 'the crying (of the people)':

(23) *I* *rongo* *au* *i* *te* [***tangi-hanga*** *hotuhotu-tanga*
T/A hear 1SG DIR.OBJ ART cry-**NMLZ** sob-NMLZ

O *ngaa taangata* *i* *te* ***tangi-hanga*** *ki* *a* *Maui Poomare*]
POSS ART people at ART funeral LOC ART M.P.
'I heard the people's sobbing at Maui Poomare's funeral.' (Bauer 1993: 48)

5.3 Summary

In sum, we have shown that in Samoan un-categorized roots can be zero-derived or overtly derived without being syntactically categorized, and that this yields non-compositional semantic interpretations. In contrast, the syntactic categorization of both un-derived and (zero) derived forms is semantically and phonologically regular. A particularly interesting finding is the 'double' status of the Samoan *-ga* suffix, which can be used for lexical as well as syntactic derivation. In the former function, it is non-categorizing and involves semantic and phonological irregularity, while in the former function it spells out a nominal functional head, with fully compositional interpretative results.

6 A Differentiated Language: Dutch

In this section we will compare our findings for Kharia, Tagalog, and Samoan to data from Dutch, in order to show how 'flexible' and 'differentiated' languages differ from one another in the way in which they derive and interpret their category-less roots. In line with the discussion on flexible languages, we will discuss two types of derivations in Dutch: those that display phonological and semantic irregularities, and those that receive fully compositional interpretations.

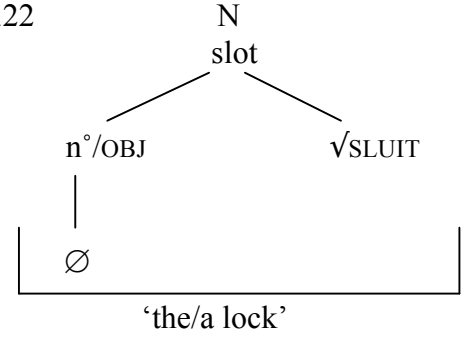
Consider first the Dutch data in (24):

(24)

	V	GLOSS	N	GLOSS
a.	drink	'to drink'	de drank	'the beverage'
b.	drink	'to drink'	de dronk	'the drink' / 'toast'
c.	bind	'to bind'	de band	'the string' / 'bandage'
d.	stinken	'to stink'	de stank	'the stench'
e.	sluit	'to lock'	het slot	'the lock'

The semantic relation between the members of these pairs is not in all cases the same, and more importantly, not in all cases predictable, as can be seen from the glosses. For example, *het slot* 'the lock' (in (24e)) is an object noun, and an interpretation of the type 'the act of locking' is not possible. Moreover, the phonology of these nouns is not fully identical to the phonology of the stem of the verbs. The semantic and phonological properties of these formations are thus similar to those of lexical, non-categorizing (zero-) derivations in the flexible languages discussed in the previous sections. The Dutch formations are crucially different, however, in that their output forms are no longer category-less: the semantic (and phonological) shift goes together with the attachment of a nominal functional head. Thus, the structure of the formations in (24), taking (24e) as an example, can be represented as in Figure 22. The notation 'n°/OBJ' is meant to capture the fact that this derivation combines a syntactic, nominalizing operation (n°), and a semantic operation (OBJ, for 'some object involved in the action denoted by the root'). Note further that the zero spell-out of the nominal head represents a simplification, since the phonological form of the output is determined by so-called readjustment rules expressing the *Ablaut*-patterns of these forms. It is an idiosyncratic property of the roots whether such related nouns exist; e.g. the verb *vind* 'to find' does not have a related nominal form #*vind*, #*vand* or #*vond*, but an irregular form *vond-st* '(a) find'. In addition, as can be seen in (24) above, some verbs, such as *drink*, have two different nominal forms, corresponding to different meanings.

Fig.22



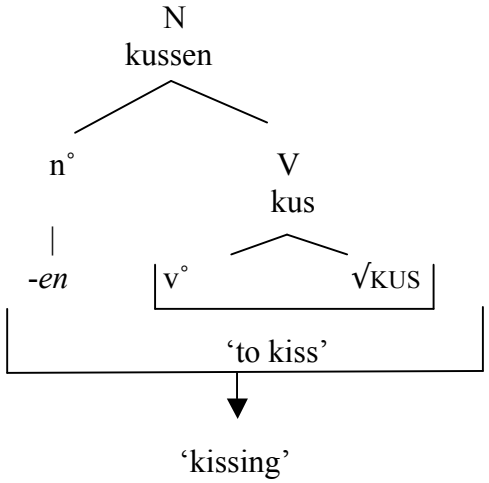
As mentioned in section 2, in differentiated languages like Dutch semantically and phonologically idiosyncratic interpretations are typically confined to the innermost domain of derivation, i.e. the combination of a category-less root with its first functional head, as it is represented in Figure 22. Once this first operation is completed, further functional heads may be attached to the resulting output form, but normally this involves no more interpretative idiosyncrasies.[17] In regard to this, consider now the Dutch formations in (25):

(25)

	V	GLOSS	N	GLOSS
a.	noem	'to name'	noemen	'naming'
b.	gooi	'to throw'	gooien	'throwing'
c.	kus	'to kiss'	kussen	'kissing'

These forms, which are often called nominalized infinitives since their phonological form is identical to infinitives, are phonologically and semantically fully regular.[18] Moreover, the process is fully productive; any Dutch verb has a nominal infinitive. The structure of the formations in (25) can be represented as in Figure 23: The nominalization is a derivation based on an already categorized form, namely the verb *kus* – itself zero-derived from a category-less root (cf. the structure in Figure 1a above) – which is combined with a functional head of a different category (noun), spelled out by the suffix *-en*, and yielding the compositional semantic interpretation '(the act of) kissing':

Fig.23



Interestingly, it seems that there are exceptions to the rule that in differentiated languages semantic and phonological idiosyncrasies arise only in the domain of the root and its first-attaching functional head. In particular, there are examples of words derived from already categorized structures that nonetheless have an idiosyncratic meaning. Consider for example the English verb *naturalize*. The morphological structure of this verb seems to be: [[[*nature*] *al*]$_A$ *ize*]$_V$. The verb is derived from the adjective *natural*. However, the adjective does not have a meaning component 'born in the country'; as one cannot refer to original inhabitants of a country as 'natural citizens'. In other words, the verb '*naturalize*' has an idiosyncratic meaning component, which is not present in the underlying categorized structure, i.e. in the underlying adjective. Along the same lines, consider the Dutch adjective *maatschappelijk* 'societal', with a morphological structure [[[*maat*]

[17] There appear to be exceptions to this generalization, however, to which we will turn shortly.
[18] For extensive discussion of these constructions see Van Haaften et al. (1985) and Zubizarreta and Van Haaften (1988), Reuland (1988).

schap]$_{N}$ *elijk*]$_{A}$, meaning 'pertaining to the society'. Unexpectedly, the meaning of the noun *maatschap* is not 'society' but 'partnership'. 'Society' translates to *maatschap-ij* in Dutch, including a suffix *–ij*. So, again we see a word with a particular idiosyncratic meaning, which does not derive compositionally from its categorized base. These examples make clear that even in differentiated languages there is no strict one-to-one correlation between syntactic categorization and semantic interpretation. This supports our analysis of flexible languages, in which semantic interpretation and syntactic categorization also constitute two separate dimensions.

7 Discussion and conclusions

The data presented in this paper show that semantic and phonological interpretation and syntactic categorization are in principle independent from one another. In differentiated languages, on the one hand, they typically go together: the semantic and phonological interpretation of a category-less root also involves its syntactic categorization. Thus, differentiated languages are *early-categorizing*: as soon as a root is derived in these languages, this derivation will involve syntactic categorization. In flexible languages, however, linguistic structures can receive semantic and phonological interpretations while remaining syntactically un-categorized. Flexible languages are thus *late-categorizing*: different types of derivations may apply that do not result in a syntactically categorized output structure (cf. Haig 2006).

In addition, we have suggested that there is a second factor, which also influences the way in which structures are interpreted: since larger, more complex structures are more easily computed than stored, such structures are more likely to receive a compositional interpretation than simplex constructions, irrespective of their being categorized or not. This may explain why in flexible languages non-categorizing derivational processes that typically operate on complex bases, such as freestanding form derivation in Kharia and gerund formation in Tagalog, receive regular semantic interpretations.

In closing, we return to the discussion surrounding semantic interpretation in flexible languages, sketched in sections 1 and 2. As we explained there, two views may be distinguished: one claiming that semantic interpretation in flexible languages should always be purely compositional (Evans and Osada 2005), and another claiming that in these languages irregular semantics are expected – if not necessarily always attested (Hengeveld and Rijkhoff 2005). Our analysis of the data in this paper makes clear that both views make sense, and that they should complement rather than exclude one another. In accordance with Evans and Osada, in flexible languages (complex combinations of) roots that are used as verbs or nouns receive compositional interpretations. This follows from our analysis of the process as a purely syntactic operation combining lexical material with functional categorial heads. However, this does not mean that idiosyncratic semantic interpretation never occurs in flexible languages. Indeed, we have shown that flexible languages exhibit non-categorizing derivational processes – both with zero and overt spell-outs – that have semantically (as well as morpho-phonologically) irregular outcomes. Crucially however, these derived structures, like simple roots, can still combine with nominal *and* verbal categorical heads, and this does not involve any further semantic increment apart from what is implied by the syntactic category itself, and the inflectional distinctions that may come along with it.

To conclude, our proposal accounts for both compositional and non-compositional derivational processes in flexible languages. As such it sheds light on the discussion about semantic shift in this type of languages. Moreover, we hope to have made a contribution to a better understanding of the notion of *flexibility*, by showing that the difference between

flexible and differentiated languages does not reside in whether or not they have distinct categories of nouns and verbs, but in how they treat their category-less roots: While in differentiated languages root-derivation means root-categorization, flexible languages can derive a root without assigning a category to it.

List of abbreviations

ACT	=	active voice	MID	=	middle voice
ART	=	article	NMLZ	=	nominalizer
CONJ	=	conjunction	OBL	=	oblique case
CV	=	conveyance voice	PL	=	plural
DIR	=	directional	PM	=	predicate marker
DIR OBJ	=	direct object marker	POSS	=	possessive
DIST	=	distal	PST	=	past tense
DUAL	=	dual	PRS	=	present tense
EMPH	=	emphatic	RDP	=	reduplication
ERG	=	ergative	SF	=	stem forming formative
GEN	=	genitive case	SG	=	singular
GENR	=	general TAM particle	SPEC	=	specifier
GER	=	gerund	STAT	=	stative
LK	=	linking element	STAT	=	stative
LOC	=	locative	T/A	=	tense-aspect
LV	=	locative voice	TAM	=	tense aspect mood

References

Aronoff, M. and Shridhar, S.N. (1987). 'Morphological levels in English and Kannada', in E. Gussman (ed.), *Rules and the lexicon*. Redakcja Wydawnictw, Katolickiego Uniwersytetu Lubelskiego, 9–22.

Bauer, W. (1993). *Maori.* London: Routledge.

Borer, H. (2003). 'Exo-skeletal versus endo-skeletal explanations: syntactic projections and the lexicon', in J. More and M. Polinsky (eds.), *The nature of explanation in linguistic theory.* Stanford: CSLI Publications, 31–67.

Croft, W. (2001). *Radical Construction Grammar.* Oxford: Oxford University Press.

--- (2005). 'Word classes, parts of speech, and syntactic argumentation', *Linguistic Typology* 9-3: 431– 441.

Embick, D. and Marantz, A. (2008). 'Architecture and Blocking', *Linguistic Inquiry* 39: 1-53.

Evans, N. and Osada, T. (2005). 'Mundari: The myth of a language without word classes', *Linguistic Typology* 9-3: 251–390.

Farrell, P. (2001). 'Functional shift as category underspecification', *English Language and Linguistics* 5-1: 109–130.

Foley, W. (1998). 'Symmetrical Voice Systems and Precategoriality in Philippine Languages', paper presented at the workshop *Voice and Grammatical Functions in Austronesian* of the LFG98 conference, Brisbane. http://www.sultry.arts.usyd.edu.au/LFG98/austro/foley/fintro.htm

Folli, R. and Harley, H. (2007). 'Causation, obligation, and argument structure: on the nature of little v', *Linguistic Inquiry* 38-2: 197–238.

Haaften, T. van, Kerke, S. van de, Middelkoop, M. and Muysken, P. (1985). 'Nominalisaties in het Nederlands', *GLOT* 8: 67–104.

Haig, G. (2006). 'Word-class distinction and morphological type: agglutinating and fusional languages reconsidered', Unpublished manuscript, University of Kiel.
Hengeveld, K. and Rijkhoff, J. (2005). 'Mundari as a flexible language', *Linguistic Typology* 9-3: 406–431.
Himmelmann, N. P. (2005). 'Tagalog', in A. Adelaar and N. P. Himmelmann (eds.), *The Austronesian Languages of Asia and Madagascar.* (Routledge Language Family Series). New York: Routledge, 350–376.
--- (2007). 'Lexical Categories and Voice in Tagalog', in P. Austin and S. Musgrave (eds.), *Voice and Grammatical Functions in Austronesian Languages.* Stanford: CSLI, 246–293.
Koptjevskaja-Tamm, M. (1993). *Nominalizations.* London: Routledge.
Marantz, A. (1997). 'No Escape from Syntax: Don't try morphological analysis in the privacy of your own lexicon', in A. Dimitriadis et al. (eds.), *Proceedings of the 21st Annual Penn Linguistics Colloquium: Penn Working Papers in Linguistics* 4: 201–225.
--- (2001). 'Words', paper presented at *West Coast Conference of Formal Linguistics*, UCLA.
Mosel, U. (2004). 'Juxtapositional constructions in Samoan', in I. Bril and F. Rivierre (eds.), *Complex predicates in Oceanic Languages.* Berlin and New York: Mouton de Gruyter, 61–294.
Mosel, U. and Hovdhaugen, E. (1992). *Samoan reference grammar* (Instituttet for sammenlignende kulturforskning, Series B: Skrifter, LXXXV). Oslo: Scandinavian University Press.
Peterson, J. (2005). 'There's a grain of truth in every "myth", or, Why the discussion of lexical classes in Mundari isn't quite over yet', *Linguistic Typology* 9-3: 39-405.
--- (2006). *Kharia. A South Munda language. Volume I: Grammatical analysis.* Habilitationsschrift, Universität Osnabrück.
--- (this volume). 'Parts of Speech in Kharia: A formal account.'
Peterson, J. and Maas, U. (2009). 'Reduplication in Kharia: the masdar as a phonologically motivated category', *Morphology* 19: 207–237.
Reuland, E. (1988). 'Relating Morphological and Syntactic Structure', in M. Everaert et al. (eds.), *Morphology and Modularity*. Dordrecht: Foris, 303–337.
Rijkhoff, J. (2008). 'On flexible and rigid nouns', *Studies in Language* 32-3: 727-752.
Schachter, P. and Otanes, F.T. (1972). *Tagalog reference grammar.* Berkeley: University of California Press.
Zubizarreta, M.L. and Haaften, T. van (1988). 'English *-ing* and Dutch *-en* nominal constructions: a case of simultaneous nominal and verbal projections', in M. Everaert et al. (eds.), *Morphology and Modularity*. Dordrecht: Foris, 361–393.

Riau Indonesian: A Language without Nouns and Verbs

David Gil

1 The English Dual

Has anybody ever written an article arguing that English does not have a grammatical category of dual number, as in, say Slovenian, Arabic, or Yimas? Almost certainly nobody has or ever will, because to do so would be pointless, not to mention with little chance of ever getting published. After all, it is self-evident that there is no dual in English, and there is no reason to believe that, contrary to appearances, all languages should, for some reason or other, have a dual. Moreover, once its non-existence is duly acknowledged, there just isn't much else to say about the English dual.

By the same token, linguists do not write articles arguing that English does not distinguish between alienable and inalienable possession, as does Dholuo; does not mark clausal switch reference, like in Diegueño; or does not have a system of multiple speech levels, as in Javanese. Similarly, for other languages, linguists do not go to great efforts trying to prove that Russian does not have definite or indefinite articles; that Turkish does not have grammatical gender; and that Sentani does not have a passive construction.

There is an obvious reason: these linguistic features, like so many others, represent properties with respect to which we take for granted that languages may vary; hence, we expect to find languages in which these features are absent. But there is another, somewhat deeper reason: all of the above examples involve a privative contrast between the presence of a feature and its absence, and in such cases, Occam's razor suggests that, as a default hypothesis, we adopt the simpler assumption that the feature is absent, until we come across positive evidence leading towards the more complex conclusion that it is present. Accordingly, while good reference grammars take note of such absences, there is no perceived need to dwell on them, or to construct explicit arguments in their justification.

However, when it comes to parts of speech, or syntactic categories, the rules of the game are somewhat different. Within many linguistic traditions, it is still assumed that all languages have the same parts-of-speech system, be it the eight parts of speech of traditional Latin grammar, or the syntactic categories posited within your favourite version of generative grammar. More recently, however, a substantial body of work, much within the framework of linguistic typology, has moved away from this assumption, and begun to explore possible patterns of cross-linguistic variation with respect to parts-of-speech systems and inventories of syntactic categories; see for example Hengeveld (1992), Vogel and Comrie eds., (2000), Croft (2001), Beck (2002), Dixon and Aikhenvald eds. (2006), and of course the present volume.

Nevertheless, even within this more recent work, there would still seem to be a lingering assumption that the default parts-of-speech system is one that is more or less similar to the familiar system of traditional Latin or generative grammar, and that any significant deviations from such systems, in particular in the direction of systems with fewer distinctions, are the ones that are somehow extraordinary and therefore worthy of special mention. One need go no further than the present volume and the term *flexibility*, featuring prominently in its title and in many of its chapters. As a reaction to the assumption that all languages are like Latin, the concern with such so-called flexible parts-of-speech systems is welcome and long-overdue. Still, the term flexibility, and much of the associated discussion, embodies an upside-down way of looking at variation in parts-of-speech systems. To be flexible means to be "able to bend or be bent repeatedly" (according to a typical dictionary definition); flexibility thus presupposes an original or canonical form, shape or configuration from which the flexible object may, under appropriate

conditions, depart. However, such a presupposition is at odds with Occam's razor, which points towards the opposite default hypothesis, namely that syntactic categories should be assumed to be absent until positive evidence is found to the effect that such distinctions are in fact present. In practical terms, what this means is that, when describing a new language, we should not ask questions such as "How do I distinguish between, say, nouns, verbs, adjectives and prepositions?", but rather questions along the lines of "What are the interesting patterns of features and how can I account for them?" — and if the best account involves positing syntactic categories such as, say, nouns, verbs, adjectives and prepositions, so be it. In theoretical terms, this suggests that preference should be given to theories in which the default inventory of syntactic categories is the minimal one, and more complex inventories characterized as more highly marked, such as, for example, Gil (2000c).

The same holds true in the particular case of concern to use here: the distinction between noun and verbs. It is commonly assumed that this is the most robust parts-of-speech distinction in language, and that if a language will make any parts-of-speech distinction at all, it will be between nouns and verbs; see for example Evans (2000: 103), Croft (2003: 183). Nevertheless, a number of recent and not-so-recent works have been devoted to arguing that various languages lack a noun/verb distinction; for example Shkarban (1992, 1995) and Gil (1993b,c, 1995) for Tagalog, Tchekhoff (1984) and Broschart (1997) for Tongan, Jelinek and Demers (1994) for Straits Salish, Swadesh (1939) for Nootka, and others. In order to dispel the notion that the noun/verb distinction is universal, such works are obviously necessary and welcome. But of course, such works are just like that hypothetical paper arguing that English has no dual. Ideally, such papers would not have to be written. Rather, linguists would describe their respective languages in the most parsimonious way, and then other linguists would examine these descriptions and make cross-linguistic generalizations about the parts-of-speech systems and syntactic category inventories that they require. But things don't work like that, and the claim that there are languages that do not distinguish between nouns and verbs remains controversial. Hence, even though it is like arguing that English has no dual, it is still necessary to go to the effort of defending the claim that some languages really do not distinguish between nouns and verbs. This, then, is the goal of the present paper, namely, to argue that in at least one language, the Riau dialect of Indonesian, there is no distinction between nouns and verbs.

In principle, the claim that Riau Indonesian does not distinguish between nouns and verbs should be supported by a simple straightforward observation to the effect that such-and-such a range of phenomena have been acounted for, all without reference to nouns and verbs, and therefore there is no need to make such a distinction. Instead the present paper adopts the following more lengthy course. Section 2 provides a brief introduction to Riau Indonesian. Section 3 discusses the notions of noun and verb and what it might mean not to have a noun/verb distinction. Section 4 examines a series of grammatical environments that in many other languages provide diagnostics for a noun/verb distinction, and shows, based on naturalistic data, that none of these criteria distinguish between nouns and verbs in Riau Indonesian. Section 5 takes a rather different tack, showing, by means of a small computational experiment, that (almost) any random combination of words in Riau Indonesian constitutes a grammatical sentence, thereby obviating the need for distinguishing between distinct open syntactic categories such as noun and verb. And Section 6 rounds up the discussion with an assessment of Riau Indonesian from a cross-linguistic typological perspective.

2 Riau Indonesian

Riau Indonesian is the variety of Malay/Indonesian spoken in informal situations by the inhabitants of Riau province in east-central Sumatra, Indonesia. The population of Riau province is linguistically and ethnically heterogeneous. Although the indigenous population is mostly Malay, a majority of the present-day inhabitants are migrants from other provinces, speaking a variety of other languages. Riau Indonesian is acquired as a native language by most or all children growing up in Riau province, whatever their ethnicity. It is the language most commonly used as a lingua franca for inter-ethnic communication, and in addition is gradually replacing other languages and dialects as a vehicle for intra-ethnic communications.

Riau Indonesian is quite different from Standard Indonesian, familiar to many general linguists from a substantial descriptive and theoretical literature. It is also distinct from the dialects spoken by the indigenous Malay community, collectively known as Riau Malay. Riau Indonesian is one of many regional varieties of colloquial Indonesian that function as lingue franche in multi-ethnic communities, of which the most well described are perhaps Jakarta Indonesian (Sneddon 2006) and Ambonese Malay (van Minde 1997). Although different from each other in numerous details, such varieties share much of their typological ground plans. In particular, the analysis of syntactic categories in Riau Indonesian presented in this paper is probably applicable to many other colloquial varieties of Indonesian, totalling tens of millions of native speakers — see Gil (2009).

The Riau Indonesian data presented in this paper are the product of several years of field work reported on in Gil (1994, 1999, 2000b, 2001a,b,c, 2002a,b, 2003b, 2004a,b, 2005a,b,c, 2006b, to appear). Most of the data presented in Section 4 makes use of a naturalistic corpus of actual utterances, either jotted down right away into a notebook or else recorded and subsequently transcribed; the remainder of the data is from the Max Planck Institute Padang Field Station Riau corpus (Gil and Litamahuputty in preparation).

3 Nouns and verbs, and what it means not to have them

In syntactic theory, a distinction is commonly made between *lexical* categories, whose members are all single words; and *phrasal* categories, whose members may consist of single words or multi-word phrases. Note that this distinction is not equipollent: lexical categories are a subtype of phrasal categories. Independently of the above, one may also distinguish between categories of the following three basic types: *morphological*, pertaining to word-internal structure; *syntactic*, pertaining to distributional properties; and *semantic*, pertaining to meaning.

How do nouns and verbs fit into the above schema? By definition, they are lexical categories; their phrasal counterparts are referred to as noun phrases and verb phrases. However, they do not fit readily into the three-way distinction between morphological, syntactic and semantic categories. Clearly, nouns and verbs cannot be defined in purely semantic terms, as words denoting things and activities respectively — witness English with action nominals such as *destruction*, or Ilgar with kinship verbs such as *ŋanimaɲyarwun* 'my father' — see Evans (2000: 103). In the description of individual languages, nouns and verbs are often defined in terms of morphological and/or syntactic properties. For example, in English, one might define a syntactic category of noun comprising those words that can occur with the definite article *the*. However, since such definitions rely crucially on language-specific constructions, they are of little use when it comes to describing the next language, or making generalizations across languages.

Following the definition provided in Gil (2000c: 197), nouns and verbs may be characterized as *semantically associated syntactic categories*, that is to say, syntactic categories that are prototypically associated with particular semantic categories. In accordance with this, nouns and verbs may be defined as follows:

(1) a. *Noun* is a lexical syntactic category prototypically associated with the semantic category of thing.
b. *Verb* is a lexical syntactic category prototypically associated with the semantic category of activity.

However, for the purposes of the present paper, we shall make use of a more general pair of definitions, relaxing the requirement that the categories in question be lexical:

(2) a. *Nominal* is a syntactic category prototypically associated with the semantic category of thing.
b. *Verbal* is a syntactic category prototypically associated with the semantic category of activity.

Thus, in accordance with the above, nouns are a subclass of nominals, and verbs a subclass of verbals; unlike nouns and verbs, nominals and verbals may consist of one word or several.

The definitions in (1) and (2) above are language-specific in that nouns and verbs, or nominals and verbals, may instantiate different syntactic categories in different languages. For example, in Gil (2000c: 194) it is suggested that in so-called "pronominal argument" languages such as Warlpiri, Lakhota and Mohawk, nominals belong to the same syntactic category, that is to say, exhibit the same distributional privileges, not as English nominals, but rather as English sentential adverbs; this is discussed further in Section 6 below. Nevertheless, in spite of their language-specific nature, the above definitions are also universal in the sense that they enable different categories to be meaningfully compared across languages. Accordingly, they are not susceptible to the kind of critique formulated by some typologists, such as Dryer (1997), Croft (2001) and Haspelmath (2007), to the effect that formal categories are cross-linguistically incommensurable.

To say that a language does not have nouns and verbs is to make the claim that the language does not have distinct lexical syntactic categories whose prototypical denotations are things and activities respectively. Or equivalently, that in the language in question, words denoting things and activities typically exhibit the same syntactic behaviour. Note, however, that a language without a noun/verb distinction might still differentiate between the phrasal categories of nominal and verbal. In such a language, words belonging to a general undifferentiated category would form the basis for distinct phrasal categories determined by their syntactic environments: give it an article and the result is a nominal phrase, add a particle expressing voice and tense/aspect/mood and the product is a verbal phrase. Indeed, analyses along these lines have been proposed for a number of languages, for example Mosel and Hovdhaugen (1992) for Samoan and Sasse (1993) for Salishan languages.

This paper makes the stronger claim that Riau Indonesian does not distinguish between nominals and verbals. This claim entails that Riau Indonesian does not distinguish between the lexical categories of nouns and verbs, but is more general. Thus, it is argued that in Riau Indonesian, there are no distinct categories of any kind (lexical or otherwise) whose prototypical denotations are things and activities respectively; or equivalently, that in Riau Indonesian, expressions (one-word or longer) denoting things and activities typically exhibit the same syntactic behaviour.

This claim that Riau Indonesian does not have a nominal/verbal distinction can be made more precise within the theory of syntactic categories proposed in Gil (2000c). Within that framework, syntactic category inventories, defined purely in terms of distributional properties, are built up recursively, in the manner of categorial grammar, from a single primitive syntactic category, S^0, by means of two category-formation operators: (a) the slash operator, familiar from other versions of categorial grammar, which takes two categories and yields a third, e.g. from categories S^0 and S^0, the derived category S^0/S^0; and (b) the kernel operator, essentially an upside-down version of X-bar theory, which takes a single category and yields another one, e.g. from category S^0 the derived category S^1. A maximally simple syntactic-category inventory would be one containing just the primitive category S^0. No language has such an inventory; however, some, including Riau Indonesian, come close. In Riau Indonesian, it is argued, there are just two syntactic categories, S^0 and S^0/S^0. The latter, S^0/S^0, is a closed class containing just a few dozen lexical items whose interpretations form a mixed bag of meanings of the sort commonly characterized as "grammatical". In contrast, almost all words, and indeed all multi-word phrases, belong to the single open syntactic category S^0. In particular, virtually all words denoting things and activities belong to the single open syntactic category S^0. In general, members of S^0 can stand on their own as complete, non-elliptical sentences, or can combine freely with other members of S^0. On the other hand, members of S^0/S^0 can only occur in construction with members of S^0, the combination of the two yielding a new S^0 expression. Thus, the goal of this paper is to demonstrate beyond reasonable doubt that in Riau Indonesian, all or almost all expressions denoting things and activities belong to the same syntactic category, S^0; that is to say, that such expressions may stand on their own as complete non-elliptical sentences, and may combine freely with any other such expressions.

4 Seeking different distributional privileges for thing and activity expressions

In languages that distinguish between nominals and verbals, the distinction is generally manifest in a number of cross-linguistically recurring grammatical environments that permit the occurrence of one to the exclusion of the other: environments in which nominals can occur but not verbals, or alternatively, ones in which verbals can occur but not nominals. Thus, if Riau Indonesian has a distinction between nominals and verbals, the chances are high that it would show up in at least some of these cross-linguistically typical environments. However, in Riau Indonesian, these grammatical environments fail to provide any evidence in support of a nominal/verbal distinction. In what follows, we shall examine a number of such environments, showing that each and every one of them may play host with equal ease to both thing and activity expressions.

4.1 Standing alone as a complete sentence

In English and some other languages, neither nominals nor verbals can stand alone as a complete sentence: if functioning as the main predicate of the sentence, they must occur in construction with an overtly expressed subject. However, as is well known, in many other languages, sometimes referred to as "pro-drop", verbs and other verbals may stand alone as complete non-elliptical sentences. What is less commonly noted is that in most such "pro-drop" languages, nouns and other nominals do not enjoy similar privileges: in predicate nominal constructions, an overt subject is still obligatory. For example, in Hebrew, *Axalti* 'I ate' is a complete non-elliptical sentence, whereas **ʕof* ‘chicken' cannot function as a compete sentence meaning 'It's a chicken'; similar facts obtain in many other languages,

including those in which nominals and verbals occur in bare form, without any agreement or other morphological markings, such as Mandarin. Thus, the ability to stand alone as a complete non-elliptical sentence provides a potential diagnostic for distinguishing between nominals and verbals.

However, in Riau Indonesian, both thing and activity expressions may stand alone as complete non-elliptical sentences, as evident in examples (3) and (4) respectively:[1]

(3) a. **Juber?**
Juber
[Group doling out popcorn at cinema, speaker suddenly wonders where his elder brother Juber is and whether he got any popcorn]
'What about Juber?'

b. **Setandannya**
one-bunch-ASSOC
[From story about a man who climbs a coconut tree to pick coconuts, throwing them to the ground where his blind son counts them according to the sound they make when they hit the ground; upon hearing a particularly loud sound he says]
'That was a whole bunch'

c. **Siapa?**
who
[Responding to knock on door]
'Who is it?'

(4) a. **Jatuh**
Fall
[Person is sitting precariously on wooden stump at end of pier, speaker warns him]
'You'll fall off'

b. **Kawin**
marry
[From story about a young man who didn't want to get married but after much coaxing, finally gave in]
'He got married'

[1] In the naturalistic examples provided in this section, the relevant thing and activity expressions are indicated in **boldface**, while — beginning with example (5) below — the forms constituting the diagnostic grammatical environments are indicated in *italics*. The actual contexts in which the examples were uttered are provided in square brackets. The interlinear glosses in this paper make use of the following abbreviations: AG agent-oriented (voice); ANIM animate (gender); ASSOC associative; AUG augmentative; DEIC deictic; DEM demonstrative; DIST distal; DISTR distributive; EP end point; EXCL exclamation; FAM familiar; FUT future; LOC locative; M masculine; NEG negative; OC object of comparison; ORD ordinal; PAT patient-oriented (voice); PERS personal; PFCT perfect; POL polarity; PROX proximal; PRPT participant; PRS present; PST past; Q question tag; REIF reifier; SG singular; TOP topic; 1 first person; 2 second person; 3 third person.

c. **Datang**
come
[Speaker watching movie for second time; an empty street is seen, and the music suggests that something is about to happen; the speaker knows that a car with the bad guys is about to appear]
'Here they come'

Whereas examples such as (4) are quite common cross-linguistically, those in (3) are somewhat more noteworthy. Of course, in many, perhaps even all languages, thing expressions may occur as complete utterances if they are licensed by a sufficiently salient context, the most obvious example being in response to a content question: *(Who is that?) John.* However, such utterances are generally judged to be structurally incomplete and semantically elliptical. In contrast, in Riau Indonesian, as suggested by the contexts in (3), thing expressions may occur as complete sentences without any kind of special contextual licensing. In particular, example (3c) represents the stereotypical response to a knock on the door. Whereas in many other languages, even "pro-drop" ones, examples corresponding to those in (3) would be unacceptable in the given contexts, in Riau Indonesian they are every bit as natural as those in (4).[2] What the parallel between (3) and (4) shows, then, is that the ability to stand alone as a complete non-elliptical sentence does not provide a diagnostic for distinguishing between nominals and verbals in Riau Indonesian. Instead, within the framework put forward in Gil (2000c), it provides the primary reason for assigning both thing and activity expressions to the single open syntactic category S^0, whose defining characteristic is that its members constitute complete sentences.

4.2 Occurring in coordination with each other

One of the most commonly cited tests for category membership is provided by coordination; in general, like terms may be coordinated while unlike terms may not. In particular, in languages with grammatically distinct nominals and verbals, these cannot be coordinated with each other. For example, in English, one cannot form the coordination *ate and clothes*; for the coordination to be grammatical, both terms must belong to the same syntactic category, e.g. *eating and clothes*, or *ate and bought clothes*. Riau Indonesian does not have a dedicated coordinator such as English *and*; however, it does have a form, *sama*, one of whose many usages corresponds to that of dedicated coordinators — see Gil (2004a,b) for detailed discussion — and may therefore be used as a diagnostic for syntactic categories.[3]

As evidenced by the following examples, in Riau Indonesian, *sama* may also be used to coordinate a thing expression with an activity expression:

---

[2] One other language allowing constructions similar to those in (3) is the geographically neighboring but otherwise quite different Singlish (Colloquial Singaporean English), discussed in Gil (2003a: 497-8).

[3] It might be argued that the many other usages of *sama* render it ineligible for diagnostics which in other languages involve dedicated coordinators. However, in at least one other language, the Austronesian Lo-Toga of the Torres Islands in Vanuatu, there is a form, *mi*, which, like Riau Indonesian *sama*, has a range of functions including 'with' and 'and': when understood as 'and', it can be used to coordinate either thing expressions or activity expressions, but crucially, not a thing expression with an activity expression (François 2010, p.c.), thereby suggesting that even macrofunctional items such as Riau Indonesian *sama* may be used as diagnostics for syntactic category membership.

(5) **Makan** *sama* **ojek**, itu aja
eat together motorcycle.taxi DEM-DEM.DIST just
[Speaker telling how he spent 5000 Rupiah in one day]
'On eating and motorcycle taxis, that's all'

(6) **Kerja** *sama* **sekolah**, gitu
work together school like- DEM-DEM.DIST
[Discussing life]
'Working and going to school, that's how it is'

In both (5) and (6) above, the first term denotes an activity while the second one denotes a thing. [4] While in the English translation, the verb must be converted into a nominal gerund before it can be coordinated with the noun, in Riau Indonesian, no such conversion is necessary, or for that matter even possible. What examples (5) and (6) show, then, is that in Riau Indonesian, coordination provides no evidence for a putative distinction between distinct nominal and verbal categories.[5]

4.3 Occurring in construction with existentials

The next eight environments to be examined all involve the coocurrence of thing and activity expressions in construction with specific forms which correspond broadly to grammatical items in other languages. The first of these to be considered is the existential marker. In English and many other languages, there is an existential construction that allows nominal expressions but not verbal ones; for example, English *There's a dog in the garden*, in contrast with **There's barked in the garden.* The latter construction makes perfect sense; however, in order to convey the desired meaning, the verbal must be nominalized: *There was barking in the garden.*

However, in Riau Indonesian, there is an existential marker *ada*. Unlike its counterparts in many other languages, *ada* has no idiosyncratic syntactic properties: it is not a "function word" but rather a "content word", belonging, together with thing and activity words, to the single open syntactic category S^0. As shown below, the existential marker *ada* may occur in construction with either thing expressions, as in (7), or activity expressions, as in (8):

[4] In Section 5.2 below, it is observed that words and larger expressions may undergo semantic type-shifting, from thing to activity and vice versa. However, there is no reason to suppose that such type shifting has occurred in the above examples, and besides, even if it had, this would be of relevance only to the semantics, not to syntactic category membership.

[5] Note that in the contexts of (5) and (6), the naturalness of the proposed English translations shows that the constraint against coordinating nominals and verbals in English and other languages is indeed a syntactic one, and not due to the kind of semantic constraints characteristic of zeugmas. Thus, if one could say *ate and motorcycle taxis* or *am working and school*, they would make perfect sense: the fact that one cannot is a purely syntactic fact about English and other similar languages.

(7) a. *Ada* **ikan** **besar** **sekali** di sana
exist fish big one-time LOC there
[Fishing]
'There's a very big fish over there'

b. Tak *ada* **dia**, Vid?
NEG exist 3 FAM-David
[Arriving at beach, looking for man who rents jet-skis]
'Isn't he here, David?'

(8) a. *Ada* **baca**?
exist read
[Speaker returning home asks interlocutor]
'Did you read the note that I left?'

b. Tak *ada* **ba-bayar**
NEG exist DISTR~pay
[At hotel, customer has ordered and paid for boat ticket from receptionist; later, receptionist gives ticket to customer and says]
'There's no additional payment'

As evident from the above examples, the semantics of the existential marker, although straightforward, is of greater generality than in many other languages. For one, as is suggested by (7b), there are no definiteness effects in Riau Indonesian: unlike in English and many other languages, definite thing expressions can occur readily in existential constructions. Next, in the examples in (8), in construction with activity expressions the existential force of *ada* produces a meaning which may perhaps be most appropriately characterized as assertive, adding emphasis to the actual occurrence of the activity — not altogether unlike English emphatic *do*.[6] Thus, as evidenced by examples (7) and (8), in Riau Indonesian, existential constructions also provide no evidence for a possible distinction between nominals and verbals.

4.4 Occurring in construction with topic markers

In many languages, topic markers can occur only with nominals, not with verbals. For example, in English, the topic marker *as for* can occur with nominals, as in *As for John, that really surprised me*, but not with verbals, as in **As for (he) hit Bill, that really surprised me*; similar observations hold with regard to topic markers in many other languages, for example Hebrew *beašer le=* and Japanese *=wa*.

In Riau Indonesian, there is a topic marker, *kalau*, which is one of the words belonging to the closed syntactic category S^0/S^0. What this means is that it cannot occur on its own as a complete sentence, but only in front of another expression belonging to the

[6] It should be noted that the semantics of *ada* in construction with activity expressions varies considerably across dialects of Malay/Indonesian. In Papuan Malay and other eastern varieties, existential *ada* in construction with activity expressions has developed an additional usage as a marker of progressive aspect. In contrast, in Jakarta Indonesian, the interpretations indicated in (8) are unavailable; instead, such constructions are generally understood as asserting the existence of one of the participants of the activity, e.g. 'Was there someone who read it?', 'Nobody pays' (such interpretations are also available, in the appropriate contexts, in Riau Indonesian). Elsewhere, a usage of the existential marker in construction with activity expressions very similar to that of Riau Indonesian is described by Thompson (1965: 216-7) for the Vietnamese existential *có*.

category S^0. However, as shown in the following examples, *kalau* can occur with either thing expressions, as in (9), or activity expressions, as in (10):

(9) a. *Kalau* **filem-filem** **sedih-sedih** **itu** tak suka
TOP DISTR~film DISTR~sad DEM-DEM.DIST NEG like
[Discussing movies]
'Those sad films, I don't like them'

b. *Kalau* **Eli** anak komplek semua kenal
TOP Eli child complex all know
[Speaker named Eli bragging about how popular he is]
'Me, all the kids in the complex know me'

(10) a. *Kalau* **kita** **mising-mising**, marah dia
TOP 1.2 DISTR~AG-noise angry 3
[About his cat]
'If one makes rustling sounds, he gets angry'

b. "*Kalau* **dapat** **burung** bawa pulang"
TOP get bird carry go.home
[From folk tale; husband going out to trap birds, wife says to him]
'"If you catch one, bring it home"'

Whereas in construction with thing expressions, as in (9), the function of *kalau* is similar to that of topic markers in other languages, in construction with activity expressions, as in (10), its usage bears a closer resemblance to that of conditional or temporal subordinators, as suggested by the English translations with 'if' (or, equally appropriately, 'when'). Nevertheless, *kalau* has a single unitary meaning of topic marker, underlying its occurrence with both thing and activity expressions. But whatever its semantic analysis, what is relevant here is that, as shown in (9) and (10), the syntactic distributional privileges of *kalau* provide no evidence for a potential nominal/verbal distinction in Riau Indonesian.

4.5 Occurring in construction with adpositions

In many languages, adpositions provide a clear diagnostic for the distinction between nominals and verbals, in that they may occur with the former but not with the latter. For example, in English, prepositions such as *from* and *for* can occur with nominals, as in *from* John and *for John* and *than John*, but not with verbals, as in **from went* and **for went*.

In Riau Indonesian, the equivalents of the above prepositions, *dari* 'from' and *untuk* 'for", are both members of the closed syntactic category S^0/S^0, occurring only in front of another expression belonging to the category S^0. Still, as evident below, these two words — in the (a) and (b) examples respectively — can occur with either thing expressions, as in (11), or activity expressions, as in (12):

(11) a. Tapi itu bukan *dari* **mami**
but DEM-DEM.DIST NEG from madam
[Taxi driver about how he gets commission from hotels and brothels]
'But not from the madam'

b. Uang yang dikasi David itu dibelikannya
money PRTP PAT-give David DEM-DEM.DIST PAT-buy-EP-ASSOC

kaset *untuk* **mamak** **dia** Vid ha
cassette for mother 3 FAM-David DEIC
[Speaking to me about his friend, whom I had given some money]
'The money that you gave him, he bought a cassette for his mother'

(12) a. Pulang istrinya tadi Vid *dari* **nyuci** kan
go.home wife-ASSOC PST.PROX FAM-David from AG-clean Q
[From story being narrated to me about a husband and his wife]
'Then his wife, who was there before, David, came home from doing the laundry, right'

b. Mana *untuk* **ngecilkan**?
which for AG-small-EP
[Learning word-processing]
'Where's the thing to make them smaller'

Similar examples can be adduced for several other similar words in Riau Indonesian corresponding to adpositions in other languages; all such words may occur with both thing and activity expressions. Thus, examples such as these show that the Riau Indonesian equivalents of adpositions provide no evidence in favour of a distinction between nominal and verbal syntactic categories.

4.6 Occurring in construction with demonstratives

In many languages, demonstratives can occur in attributive construction with nominals but not verbals, as can be illustrated in the contrast between **this chicken* and **this ate*.

In Riau Indonesian, there are two demonstratives, proximal *ni* and distal *tu*, which may occur either on their own, or as in the examples below with the additional demonstrative prefix *i-*. These forms are members of the single open syntactic category S^0, and may thus combine freely with other members of S^0, the resulting constructions being of either attributive or predicative nature — there is no clear contrast between the two in Riau Indonesian. And, as shown below, these forms may occur with either thing expressions, as in (13), or activity expressions, as in (14):

(13) a. O, nunggu **abang** *ini*
EXCL AG-wait elder.brother DEM-DEM.PROX
[In shared taxi waiting to fill up, man enters taxi and driver starts the engine, at which point another passenger realizes driver was waiting for this man and comments]
'Oh, waiting for him'

b. Mana **orang** *itu*?
which person DEM-DEM.DIST
[Group on motorcycle taxis; my driver loses sight of the rest, and asks]
'Where are the others?'

(14) a. Aku **mau** **terkencing** *ini*
1SG want NON.AG-urinate DEM-DEM.PROX
[Banging on door, wanting to enter room]
'I want to pee'

b. Apa **kalian** **mau** **li-lihat** *itu?*
what 2PL want DISTR~see DEM-DEM.DIST
[Crowd of curious children gathered around foreigner, woman addresses them]
'What are you all looking at?'

As suggested by the above examples, Riau Indonesian demonstratives also fail to provide any reason for distinguishing between putative syntactic categories of nominals and verbals.

4.7 Occurring in construction with distributive universal quantifiers

In many languages, distributive universal quantifiers can occur in direct construction with nominals but not verbals. For example, in English, the distributive universal quantifier *every* can quanfity nouns, as in *every student*, but not verbs, as in **every came*; in order to quantify over activities or events, a more complex construction is called for, one in which the quantified-over entities are expressed with the noun *time*, as in *every time they came*. Similar asymmetries between nominal and verbal quantification are present in many other languages; see Gil (1993a) for examples and discussion.

In Riau Indonesian, the distributive universal quantifier is *tiap*, often occurring in reduplicated form, *tiap-tiap*, or with the prefix *se-* 'one' yielding the form *setiap*. These various forms of the quantifier have similar meaning and also similar syntactic behaviour; as members of the closed syntactic category S^0/S^0, they can only occur in front of another expression belonging to the category S^0. As shown below, the distributive universal quantifiers may occur in construction with either thing expressions, as in (15), or activity expressions, as in (16):

(15) Itu *tiap-tiap* **desa** ada
DEM-DEM.DIST DISTR~every village exist
[In response to question about a rural military office]
'Every village has one'

(16) *Setiap* **dia** **datang** **di** **diskotik** dia kasi sepuluh ribu
one-every 3 come LOC discotheque 3 give one-ten thousand
dua puluh ribu
two ten thousand
'Every time he comes to the discotheque, he gives ten thousand or twenty thousand'
[About a Japanese tourist]

While in (15) *tiap-tiap* quantifies over things, in (16) it quantifies over activities or events. Although Riau Indonesian offers a possible paraphrase of (16) containing the word *kali* 'time", namely *Setiap kali dia datang di diskotik ...,* the presence of *kali* when quantifying

over activities or events is optional. (When present, the effect of *kali* is of a semantic classificatory nature resembling that of the numeral classifiers which are similarly optional in the case of quantification over things.) Thus, by occurring in direct construction with either thing or activity expressions, the distributive universal quantifier *tiap* and its variants also provide no support for positing distinct nominal and verbal syntactic categories in Riau Indonesian.

4.8 Occurring in construction with tense/aspect markers

Whereas the preceding five subsections examined constructions which, in other languages, are characteristically nominal, this subsection takes a look at forms which, cross-linguistically, typically occur with verbals rather than nominals, namely tense and aspect markers.

Riau Indonesian has no grammatical tense or aspect; that is to say, it has no tense or aspect markers whose occurrence is obligatory and whose grammatical properties are unique or otherwise idiosyncratic. What it does have, however, is a small number of words with abstract meanings corresponding closely to those associated with tense and aspect markers in other languages, but with distributional privileges that are identical to almost all other words and longer expressions in the language. These words are thus members of the single open syntactic category S^0: they occur freely on their own as complete non-elliptical sentences, and they combine freely with other members of S^0. Two such words are *tadi*, expressing proximal past time, and *udah* (and its variants *sudah* and *dah*), whose meaning lies somewhere in the range between perfective and perfect. As shown in (17) and (18), *tadi* may occur in construction with either thing or activity expressions; similarly, as exemplified in (19) and (20), *udah* may also occur in construction with expressions belonging to either of these two semantic classes:

(17) a. Mana **Keling** *tadi*?
which Keling PST.PROX
[Wondering about his friend who was present a short while before]
'Where's Keling?'

b. **Penumpang** *tadi* masuk sini
passenger PST.PROX enter LOC-DEM-DEM.PROX
[Aboard boat, petrol runs out and boat stops in mid-ocean; another boat arrives from opposite direction; our boatmen stop the other boat, take its passengers on board, and one of our crew then takes the other boat to a nearby port to buy petrol. Later, the other boat returns with the petrol, and one of its boatmen calls out for its passengers to board it again.]
'The passengers from before go in here'

(18) a. Aku **masuk** **terus** *tadi*
1SG enter continue PST.PROX
[About a billiards game he just played]
'I just kept on knocking them in'

b. Apa **dibilang** *tadi* kau?
what PAT-say PST.PROX 2SG
[Quizzing friend about phonecall he just made]
'What did he say to you?'

(19) a. *Udah* **ongkos** taksinya?
PFCT fare taxi-ASSOC
[Getting out of taxi at host's house, host asks]
'Have you paid the taxi yet?'

b. Yang tadi suaranya *sudah* **perempuan**
PRTP PST.PROX voice-ASSOC PFCT woman
[At transvestite bar, commenting on another person who was present a short while ago]
'The one just before, her voice was already that of a woman'

(20) a. *Dah* **tak** **tahan** dia 'kan, Vid
PFCT NEG withstand 3 Q FAM-David
[From story about a peeping Tom; he gets aroused]
'He couldn't restrain himself any longer, David'

b. Anak tukang nyemir *udah* **bayar**?
child AG AG-polish PFCT pay
[Group of people drinking in food court; having bought drinks for some shoeshine boys, one asks another]
'Have the shoeshine boy's drinks been paid for yet?'

Whereas in (18) *tadi* specifies the time of an activity, in (17), the same form specifies a time contextually associated with a thing. In (17a), the speaker is not entirely certain that the hearer is familiar with the name *Keling*,[7] so to further ensure referential success, he adds the qualifier *tadi*, in recognition of the fact that Keling was together with both speaker and hearer just a short while ago. Similarly, in (17b), the boatman wishes to address the passengers from his boat to the exclusion of the passengers from the stranded boat, and the easiest way of doing this in the given context is to qualify the thing word *penumpang* 'passenger' with *tadi*, which in this case refers to the moment in which the passengers moved from his boat to the stranded one.[8] In a somewhat different pattern, whereas in (20) *udah* imposes aspectual structure on an activity expressed by an activity word, in (19) it imposes the same structure on an implicit activity associated with an overtly expressed thing word: the activity of paying with the thing word *ongkos* 'fare', and the activity of becoming with the thing word *perempuan* 'woman'. (Such semantic type shifts are discussed further in Section 5.2 below.) What it at issue here though is not semantics but rather syntax and distributional privileges: syntactically, it is clear that *udah* goes with activity expressions in (20) but with thing expressions in (19). Thus, as shown by the above examples, the closest that Riau Indonesian can come to tense and aspect markers also fails to provide any evidence for distinguishing between syntactic categories of nominals and verbals.

[7] Actually, *Keling* is a nickname commonly given to people with dark complexion (its origin is in the name of the ancient Kalinga kingdom of India, whose people were considered to be dark-skinned).
[8] In both examples in (17), in a different context, the same string of words would allow an alternative interpretation in which *tadi* is understood as qualifying an activity: 'Where was Keling just before' and 'The passengers went in here just before'. However, such alternative interpretations would be associated with different constituencies, in which *tadi* is not grouped together with the thing word, as it is in (17); such different constituencies are characteristically reflected by different intonational parsings.

4.9 Occurring in construction with *yang*

To this point we have examined eight potential diagnostics for distinguishing between nominals and verbals that were motivated by universal cross-linguistic considerations, that is to say, constructions that in other languages are known to differentiate between nominal and verbal categories. We shall now examine two additional diagnostics that have been proposed specifically for Malay/Indonesian, making reference to specific forms in Malay/Indonesian.

The first of these pertains to the marker *yang*. In most descriptions of Standard Malay and Indonesian, *yang* is described as a relativizer, occurring in front of a relative clause, generally headed by a verbal expression. Similar descriptions are also offered for *yang* in varieties of colloquial Malay/Indonesian, for example Tjung (2006) for Jakarta Indonesian. Interestingly, prescriptive and pedagogical grammars of Malay and Indonesian are often explicit about the ungrammaticality of *yang* in construction with nouns, rather than verbs, for example Mintz (1994: 51). Nevertheless, examples of *yang* occurring in construction with a nominal expression are provided by Azhar (1988: 143-8) for Classical Malay and by Sneddon (1996: 287-8) for Standard Indonesian.

In Riau Indonesian, the form *yang* also occurs, in constructions bearing a superficial resemblance to their counterparts in Standard Malay/Indonesian, and also in lots of constructions that look quite different. The marker *yang* belongs to the closed syntactic category S^0/S^0, occurring only in front of another expression belonging to the category S^0. Thus, as shown below, it may occur in construction either with thing expressions, as in (21), or with activity expressions, as in (22):

(21) a. Dia mau bunuh *yang* **perempuan**?
3 want kill PRTP woman
[Watching movie]
'Is he going to kill the woman?'

b. *Yang* **duri** keluar sendiri
PRTP thorn go.out one-AG-stand
[Streetside medicine man describing the effects of an ointment]
'The thorn will come out on its own'

(22) a. Ini *yang* **beliin**
DEM-DEM.PROX PRTP buy-EP
[Speaker asks friend for money for fishball soup; friend observes that he's already eating some; speaker points to other friend sitting next to him and says]
'He's the one who bought it for me'

b. Dia *yang* **nyuci**?
3 PRTP AG-clean
[Asking about hotel roomboy]
'Is he the one who does the laundry?'

Whereas the examples in (22) resemble their counterparts in Standard Malay/Indonesian and are amenable to relative-clause translations into English, the examples in (21) are the sort of constructions that are ruled out by the prescriptivists — with *yang* occurring in construction with a thing expression. The indifference of *yang* to the semantic category of the expression it precedes is illustrated vividly in the following example, where *yang* precedes a thing word and right after that an activity expression:

(23) Ini *yang* **bas** *yang* **kita** **naik** **dulu**
DEM-DEM.PROX PRTP bus PRTP 1.2 ride previous
[Catching sight of bus]
'That's the bus we took before'

Thus, contrary to the prescriptive grammarians, in Riau Indonesian at least, the marker *yang* provides no evidence whatsoever for a distinction between nominal and verbal categories.

Examples such as the above raise the question what kind of a creature the marker *yang* actually is. Since there is no evidence to the effect that expressions occurring in construction with *yang* contain gaps or empty positions, *yang* is not a relativizer in the usual sense of the word.[9] Rather, *yang* can be characterized in purely semantic terms, with reference to the semantic frame of the expression to which it applies. Specifically, given an expression E with meaning M, the derived expression *yang E* is interpreted as having the meaning PRTP (M), or 'participant belonging to the semantic frame of M'. This can be seen most clearly in those cases where *yang* occurs in construction with an activity expression. For example, in (22a) and (22b) *yang beliin* and *yang nyuci* refer to the agent of 'buy' and 'clean' respectively, while in the second part of (23), *yang kita naik dulu* refers to the patient of 'ride'. However, thing words also have semantic frames, and hence *yang* may occur in construction with such expressions too. In order to make sense of this, it is necessary to enrich the traditional inventory of thematic roles with an additional role, namely, *essant*. The essant role can be thought of as that characteristic of the subject of symmetric predications of the kind that, in many languages, make use of a copula. For example, in English, the demonstrative *this* bears the essant role in constructions such as *This is John*, *This is a student*, *This is a murder*.[10] In (21a), (21b), and the first part of (23), the expressions *yang perempuan*, *yang duri* and *yang bas* refer to the essants of 'woman', 'thorn' and 'bus', which of course are also a woman, a thorn and a bus respectively. The difference in meaning between bare thing expressions and their counterparts with *yang* is subtle, and can only be fully appreciated through examination of the discourse function of *yang* in such constructions. But this need not concern us here. For present purposes it suffices to point out that Riau Indonesian *yang* may occur in construction with either thing

---

[9] Talking about *yang* and empty positions in Malay/Indonesian is a bit like talking about nouns and verbs, cf. Section 1. In principle, a simple statement to the effect that there is no evidence for empty positions should end the discussion. But since many constructions with *yang* get translated into English relative clauses, which have gaps, it is commonly assumed that the Malay/Indonesian constructions with *yang* must also contain empty positions, and if one wishes to claim otherwise, one is expected to muster explicit arguments. Instead, however, the absence of gaps should be the default hypothesis, for constructions with *yang* just like anywhere else in the grammar, and the burden of the proof should be on those who wish to show that such empty positions do indeed exist. In other languages, an important source of evidence for gaps in relative clause constructions involves island constraints; however, in Gil (2000a) it is argued that, in Riau Indonesian, at least, the corresponding range of phenomena is amenable to a semantic analysis not involving gaps.

[10] Note that whereas the first construction, *This is John*, is equative, the latter two, *This is a student* and *This is murder*, involve the predication of class inclusion. The role of essant is thus more general than Carnie's (1997: 62) "attribute recipient", which is limited in its application to equative constructions. However, the role of essant is less general than the traditional notion of theme, which would also apply to *this* in constructions such as *This is interesting*. Note also that whereas in the first two constructions, the essant is a thing expression, in the third it is an activity expression, the demonstrative here denoting the activity of murder. The crucial feature of all three constructions, to the exclusion of *This is interesting*, is that of semantic symmetry, both terms, essant and predicate, belonging to the same semantic category. Motivation for the delimitation of the essant role is provided by the differential distribution of negative markers discussed in Section 4.10 below.

or activity expressions, and hence its proper analysis makes no reference to nominal and verbal syntactic categories.

4.10 Occurring in construction with different negative markers

When confronted with the claim that Riau Indonesian does not distinguish between nominals and verbals, one of the most common reactions from scholars of Malay/Indonesian is to ask: What about the two negative markers, *bukan* and *tidak*? Indeed, many descriptions of the standard languages claim that *bukan* occurs with nouns while *tidak* occurs with verbs; see, for example, Macdonald and Soenjono (1967: 160-1) and Sneddon (1966: 195-7) for Standard Indonesian, and Nik Safiah Karim (1996: 249) for Standard Malay. This claim has even found its way into the general typological literature, for example Stassen (1997: 48). However, even within the standard languages, the story is not cut and dry; in fact, most grammatical descriptions acknowledge exceptions to the noun/verb generalization, which they attempt to account for semantically, most commonly by appealing to an additional "contrastive" feature associated with the would-be nominal negator *bukan*, supposedly licensing its occurrence in construction with verbs.

Many colloquial varieties of Malay/Indonesian preserve the distinction between *bukan* and *tidak*, though the actual forms, most often of the second member of the pair, may vary. For example, Jakarta Indonesian contrasts *bukan* with *nggak, kagak* or *gak*, claimed by Sneddon (2006: 56-7) to occur in construction with nouns and verbs respectively. Similarly, Palembang Malay contrasts *bukan* with *tak*, described by Sainul et al (1987: 166-7) as being associated with nouns and verbs respectively. And in a somewhat more complex pattern, Sri Lanka Malay contrasts *bukang* with either *thera-* or *thama-*, argued by Nordhoff (this volume) to occur with nouns and verbs respectively, the choice of verbal form being further conditioned by the tense of the verb.

In Riau Indonesian, *bukan* contrasts most commonly with *tak*, though *nggak* and *ndak* are also available as alternative forms. Both *bukan* and *tak* (and its variants) belong to the single open syntactic category S^0, occurring freely on their own and in combination with other members of S^0. The following examples illustrate the usage of *bukan* in (24) and (25), and of *tak* in (26) and (27). As is evident in these examples, both *bukan* and *tak* occur freely in construction with either thing expressions, as in (24) and (26), or activity expressions, as in (25) and (27):

(24) a. *Bukan* **mercun** bunga api David beli
NEG firecracker flower fire David buy
[Complaining about my holiday presents]
'What you bought wasn't firecrackers but fireworks'

b. **Uri** *bukan*
Uri NEG
[Speaker hears voices of two people in other room speaking English and recognizes one of them, but the other voice is of somebody he does not know; wondering who it might be, he thinks of the other foreigner who he knows, whose name is Uri, but then realizes that the voice is not his]
'Uri it isn't'

(25) a. *Bukan* **be-beli** **apa-apa**, nengok aja
NEG DISTR~buy DISTR~what AG-look just
[Speaker trying to convince friend to go with him to shopping mall]
'What I want to do isn't to buy anything, but just to look around'

b. Ini *bukan* **buka**
DEM-DEM.PROX NEG open
[Speaker taking apart a complicated electronic device; interlocutor complains, suggesting he shouldn't be doing that; speaker justifies his activity]
'This isn't taking anything apart'

(26) a. Dia *tak* **sekolah**
3 NEG school
[Child talking about friend]
'He doesn't go to school'

b. Ini *tak* **pameran** do
DEM-DEM.PROX NEG exhibit-AUG NEG.POL
[At coffeeshop table surrounded by group of onlookers; another group arrives, but one of the original group members tries to fend them off]
'We're not putting on an exhibition here'

(27) a. *Tak* **masuk** do
NEG enter NEG.POL
[Playing billiards, having just missed shot]
'It didn't go in'

b. Tahu aku *tak* **beli**
know 1SG NEG buy
[About some ice cream which turned out not to be nice]
' If I had known, I wouldn't have bought it '

Thus, as clearly shown by the above examples, the choice between *bukan* and *tak* has nothing to do with putative nominal and verbal syntactic categories.

What, then, governs the choice between these different negators? As suggested for Standard Malay/Indonesian, in Riau Indonesian too, *bukan* is often endowed with contrastive force, with an alternative to the negated entity either implied by the context or in some cases overtly expressed. For example, in (24a), *bukan* negates *mercun* 'firecrackers', contrasting it with *bunga api* 'fireworks', while in (25a), *bukan* negates *be-beli apa-apa* 'buy anything', contrasting it with *nengok* 'look around'. However, in many other instances, *bukan* does not seem to be contrastive. For example, in (24b), *bukan* negates *Uri* 'Uri', but the speaker has no idea who the voice he hears might belong to; similarly, in (25b), *bukan* negates *buka* 'take apart', but the speaker is not offering any alternative description of what he is actually doing. Thus, contrastive force is not a necessary property of negation with *bukan*. Nor for that matter is it an unavailable property of negation with *tak*. Although none of the examples in (26) and (27) are contrastive, is quite easy to modify them in order to render them as such. For example, (26a) could be continued with *...dia main-main aja* '... he just plays around', resulting in a contrastive interpretation for the negator *tak*. Thus, although *bukan* seems to lend itself more frequently than *tak* to contrastive interpretations, this is not an inherent property of the distinction between these two negators, and therefore it cannot constitute the basis for an account of their different distributions.

Instead, the distinction between *bukan* and *tak* is a semantic one, making reference to the semantic frame of the negated expression, and, in particular, the thematic role of essant, introduced in the preceding subsection. Specifically, *bukan* negates an expression in relationship to its essant, whereas *tak* negates an expression with respect to any of its other thematic roles. In the above examples, this distinction is readily evident in the English translations: those of the sentences with *bukan*, in (24) and (25), all contain a negated form of the copula 'be', which, in English, is characteristic of constructions expressing the essant relationship. Whereas in (24) the essant is that of a thing expression, in (25) it is that of an activity expression; but in both cases, the expression is being negated with respect to its essant, and hence the negation is with *bukan*.[11]

While the semantic frame of thing expressions typically has essant as the only thematic role, the semantic frame of activity expressions is generally richer, with essant occurring alongside other thematic roles such as agent, patient, and so forth. Accordingly, it is with activity expressions that the negator *tak* occurs most readily, as in (27). In order for *tak* to occur with thing expressions, as in (26), such expressions must undergo semantic conversion, in order to be understood as activities (the notion of semantic conversion is further discussed in Section 5.2 below). Thus, in (26a), *sekolah*, whose basic meaning is 'school', undergoes type shift and is understood as 'go to school'; similarly, in (26b), *pameran*, whose basic meaning is 'exhibition', undergoes type shift and is understood as 'put on an exhibition'. Note that while the use of *sekolah* to denote the activity of going to school is a conventionalized and high-frequency meaning extension, the use of *pameran* to mean 'put on an exhibition' is an example of a creative one-off type shift. Crucially, though, the type shift in examples such as these is semantic, not syntactic.

Thus, the distinction between *bukan* and *tak* is itself entirely semantic, making reference only to the thematic roles associated with the negated expression. Because of the different semantic frames of thing and activity expressions, *bukan* tends to occur more often with thing expressions, while *tak* occurs with much greater frequency with activity expressions. Nevertheless, both *bukan* and *tak* have the potential of occurring with either thing or activity expressions, as evidenced in (24) - (27) above; accordingly, the contrast between these two negators provides no evidence whatsoever in support of distinct syntactic categories associated with these semantic types, namely, nominals and verbals.

4.11 Interim summary

In the preceding ten subsections, we examined ten syntactic environments which were deemed likely to yield a distinction between nominal and verbal syntactic categories: the first eight motivated by cross-linguistic considerations, the latter two by claims specific to

---

[11] The function of *bukan* when occurring in construction with activity expressions can be highlighted by contrasting example (25b) with its hypothetical variant in which *bukan* is replaced with *tak*:

(i) Ini tak buka
DEM-DEM.PROX NEG open
'This (thing) isn't open' / 'This (person) isn't opening it'

Since in (25b) *bukan* negates *buka* in relation to its essant role, it is this role that is associated with the demonstrative *ini*, which accordingly refers to the activity of taking something apart. In contrast, in the variant in (i), *tak* negates *buka* in regard to some other role, of which the most readily available one is theme; accordingly, the most obvious interpretation of (i) is one in which the demonstrative *ini* denotes the thing that is not open. However, in other contexts, the role with respect to which *buka* is negated can be different, for example, that of agent, in which case the demonstrative *ini* would be understood as referring to the person who is not engaged in the activity of opening.

Malay/Indonesian. For each of these ten environments, it turns out that there is no difference in the syntactic behaviour of thing and activity expressions. Accordingly, these ten potential diagnostics provide no evidence in favour of a nominal/verbal distinction in Riau Indonesian. Of course, this does not prove that there is no such distinction: it could be there, waiting to be discovered, in some other syntactic environment yet to be examined. But until it is, the most appropriate conclusion to be drawn is that Riau Indonesian does not distinguish between nominal and verbal syntactic categories.

5 Bidirectionality, Compositionality and Exhaustiveness

In an important recent contribution, Evans and Osada (2005) propose three criteria which they maintain must be met before a language can appropriately be characterized as lacking a noun/verb distinction: *bidirectionality*, *compositionality*, and *exhaustiveness*. While open-minded with regard to whether such languages do actually exist, they argue that many previous claims that particular languages lack a noun/verb distinction fail to meet these criteria, and therefore must be rejected. Their argument is illustrated with reference to claims made with regard to the Austroasiatic language Mundari. Whatever the specific merits of Evans and Osada's (re)analysis of Mundari, we shall now show that Riau Indonesian meets their three criteria, and hence may reasonably be said to lack a noun/verb or nominal/verbal distinction.

5.1 Bidirectionality

The first criterion, a fuller label for which is bidirectional distributional equivalence, specifies that all thing and activity words must enjoy the same distributional privileges. This is essentially the point that was demonstrated, at some length, in Section 4 above. However, as pointed out by Evans and Osada, such claims often fail to meet the condition of bidirectionality, namely that the equivalence work in both directions. For example, in Mundari and in Nootka, nouns occur freely in verbal positions, however, in order for verbs to occur in nominal positions, some additional morphosyntactic adjustment needs to be made, such as the introduction of a determiner. Accordingly, Evans and Osada argue that the monocategorial analyses proposed for these two languages ought to be rejected.

However, Riau Indonesian clearly does meet the condition of bidirectionality. Reviewing the diagnostics discussed in Section 4, those in 4.1 and 4.3 - 4.7 involve activity expressions occurring in constructions where one might have expected to find a thing expression (precisely the situation which, according to Evans and Osada, does not obtain in Nootka and Mundari), while those in 4.8 and 4.9 instantiate the mirror image state of affairs in which thing expressions occur in environments where an activity expression would be expected; that in 4.10 exhibits both possibilities (each associated with a different negator). Thus, in Riau Indonesian, thing expressions can be replaced with activity expressions and activity expressions with thing expressions freely and without any need for further morphosyntactic adjustment — there are no morphosyntactic nominal positions or verbal positions to speak of. This, then, is what Osada and Evans call bidirectional distributional equivalence, though the term bidirectional is a bit of a misnomer, since it implies that something is going somewhere, which in a true monocategorial language is not the case: without characteristic nominal and verbal environments, there is no morphosyntactic conversion, and everything is where it belongs.

5.2 Compositionality

The criterion of compositionality states that the meaning of an expression must be predictable from the meanings of its parts plus the meanings associated with its morphosyntactic environment. The absence of compositionality in English noun-to-verb conversions is manifest in the idiosyncratic differences in meaning between verbs such as *shovel, cup, can,* and *spearhead*, which, although all formed from nouns denoting instruments, are associated with quite variegated meanings: 'use X as instrument', 'form into shape of X', 'place in X', and 'act like an X' respectively. It is such irregularities, Evans and Osada argue, that preclude the characterization of English as a language without a noun/verb distinction.

Since this criterion is a semantic one, it is not clear to what extent it is relevant to the determination of syntactic categories. Still, it is worth examining the Riau Indonesian facts in its light. In Riau Indonesian, like in English and presumably all languages, words are prototypically associated with specific semantic categories such as thing and activity. At first blush, it would appear as though Riau Indonesian fails to meet the criterion of compositionality; examples can be adduced which bear a close resemblance to Evans and Osada's cases of semantically non-predictable conversions in English, Mundari and other languages. One such example was encountered in the previous section: the conventionalized type shift of *sekolah* from thing word 'school' to activity word 'go to school' in (26a). Going in the other direction, *makan* 'eat' and *tembak* 'shoot' are two words whose prototypical meanings are activities containing an agent and a patient in their semantic frames. Both words, though, are commonly and conventionally used also to denote things, but with idiosyncratic and unpredictable meanings: *makan* as 'food', the patient of *makan* 'eat', but *tembak* as 'gun', the instrument of *tembak* 'shoot'.[12] In contrast, other semantically similar words, such as *kejar* 'chase' and *bilang* 'say', have no corresponding conventionalized thing interpretations. So does Riau Indonesian work the same way as English, then? Not at all.

To begin, in English, words such as *shovel, cup, can,* and *spearhead* exhibit different grammatical behaviour depending on whether they have undergone semantic type shift: while *shovel* the thing behaves like a noun, *shovel* the activity displays the grammatical properties of a verb. In contrast, in Riau Indonesian, words such as *sekolah*, *makan* and *tembak* exhibit the same grammatical behaviour regardless of whether they have undergone semantic type shift: *sekolah* 'school' behaves just like *sekolah* 'go to school'. Thus, although for both languages such conventionalized semantic type shifts must be represented in their respective lexicons, in Riau Indonesian such lexical information has nothing to do with syntactic categories and a would-be noun/verb distinction.

But there is an even more important difference between the two languages. In English, words such as *shovel, cup, can* and *spearhead*, however numerous, stand out against the default situation of nouns being nouns, and not being able to undergo zero conversion: to the best of my knowledge there are no verbs *knob, stool, spatula*, or *computer*, under any kind of interpretation. In contrast, in Riau Indonesian, all words would seem to have the potential for undergoing zero-marked semantic conversion from

[12] It is interesting, and probably no coincidence, that the same word, *makan*, with the same semantic conversion from 'eat' to 'food', is cited by Ewing (2005), and then cited in turn by Nordhoff (this volume), as evidence for syntactic-category flexibility in the colloquial Indonesian of Java. Evans and Osada would presumably reject this argument as it stands, on the grounds that it violates compositionality; indeed, they would also be able to point out that the same conversion from 'eat' to 'food' is found in the word *jom* in Mundari.

their primary semantic category, for example thing or activity, to any other semantic category; moreover, such type shift is semantically unconstrained — anything goes, provided it makes sense in the given linguistic and extra-linguistic context. Some examples of creative type-shifting from thig to activity were encountered in Section 4 above: *ongkos* 'fare' > 'pay the fare' in (19a), *perempuan* 'woman' > 'become a woman' in (19b), and *pameran* 'exhibition' > 'put on an exhibition' in (26b). Going in the other direction, any expression whose prototypical meaning is that of an activity with one or more participants in its semantic frame can undergo a zero-marked semantic conversion, resulting in a thing interpretation where the thing in question is one of the participants, or any other entity associated in some way with the activity. Moreover, this is true not only of words such as *kejar* 'chase' and *bilang* 'say', but also of words such as *makan* 'eat' and *tembak* 'shoot', the only difference being that for the latter, the regular compositional meanings are less salient than the idiosyncratic conventionalized lexical meanings mentioned previously. In fact, type shifting of the kind illustrated above is not limited to individual words but can apply also to longer expressions: it thus applies to syntactic representations, not to lexical ones.[13]

In summary, then, English and Riau Indonesian are alike in that they both have cases of conventionalized, semantically-unpredictable zero-conversion of words from one semantic category to another, whose most appropriate locus of representation is thus in the lexicon. However, they are diametrically opposed with respect to what happens elsewhere in the syntax. Whereas in English, there is no zero-marked semantic conversion outside of the lexicon, in Riau Indonesian, zero-marked type shifting is possible everywhere, for all words and larger expressions, and without arbitrary constraints on the nature of the resulting interpretation. Thus, Riau Indonesian may be said to meet the criterion of compositionality.

Semantic conversion of this kind can be represented with the *Association Operator*, introduced in Gil (2005b,c). Given an expression E with meaning M, application of the Association Operator A yields a derived meaning A (M), or 'entity associated with M'. Cross-linguistically, the most common manifestation of the Association Operator is in genitive constructions; for example, in English, the meaning of *John's* is A (JOHN), or 'entity associated with John'. However, the Association Operator may also apply in the absence of overt morphosyntactic marking, the most well-known case of this involving the range of phenomena commonly subsumed under the label of metonymy. In general, though, genitival and metonymical constructions do not necessarily involve type shifting.

---

[13] Some examples of the zero-marked semantic conversion of multi-word expressions are provided by the bold-faced expressions in the following utterances:

(i) Mana **aku beli tadi**
which 1SG buy PST.PROX
[Looking for a tube of toothpaste he had bought a short while before]
'Where's the thing I bought just before?'

(ii) Kalau saya **minuman yang ada gas**, sakit perut saya
TOP 1SG drink-AUG PRPT exist gas hurt stomach 1SG
[Explaining his turning down the offer of a drink]
'If I have drinks with gas, my stomach hurts'

In (i), the activity-denoting expression *aku beli tadi* 'I bought just before' undergoes type-shift and comes to denote a thing 'the thing I bought just before'; this kind of construction is discussed in more detail in Gil (1994, 2000b). Conversely, in (ii), the thing-denoting expression *minuman yang ada gas* 'drinks with gas' undergoes type-shifting and is understood as referring to an activity 'have drinks with gas'. (Note that the head of the construction, *minuman*, is itself the product of an idiosyncratic lexically-conditioned typeshift of the kind discussed previously, in which the activity word *minum* 'drink (imbibe)' combines with the suffix *-an* to produce the thing word 'drink (beverage)'.)

In Riau Indonesian, however, application of the Association Operator is much more pervasive, applying across the board in the absence of any overt morphosyntactic marking, and resulting in rampant type shifting across major semantic categories such as thing and activity. For example, the three cases of conversion in examples (19a), (19b) and (26b) may be represented as A (FARE), A (WOMAN) and A (EXHIBITION) respectively. By containing no more than what is available from the meanings of their constituent parts, such representations of type shifting underscore how semantic conversion in Riau Indonesian satisfies the condition of compositionality.

5.3 Exhaustiveness

The criterion of exhaustiveness embodies a higher-level methodological principle which governs, in turn, the application of the two preceding criteria; it states, simply, that bidirectional distributional equivalence and compositionality must apply without exception to all items, and not to just a subset of examples judiciously chosen in order to make a certain point. This should be obvious, except that, as Evans and Osada point out, many previous claims to the effect that a language lacks a noun/verb distinction make reference to a relatively small number of words or expressions, raising the suspicion that a larger sample of words would result in a somewhat different picture of reality.

In principle, exhaustiveness can only be tested by actually examining each and every one of the items under investigation. In practice, however, this is not feasible. With perhaps hundreds of thousands of words each potentially occurring in any number of diagnostic environments, the number of questions that need to be asked is, by orders of magnitude, beyond the collective capacity of any researcher working with any number of cooperative native speakers. Searching an electronic corpus is no solution, either, because, invariably, there will be plenty of accidental gaps, where grammatical combinations of uncommon words just happen not to show up in the corpus. Instead, the most practical way to test exhaustiveness is by random sampling; this is the method that is adopted here.

The claim that is tested here is that all members of the syntactic category S^0 have the same syntactic distribution. This claim is somewhat stronger than the claim that Riau Indonesian has no nominal/verbal distinction, since it is concerned with the distribution not just of thing and activity words but also of words belonging to other semantic categories such as property, place, time, and so forth. In order to test this claim, the following procedure was adopted. First, a random list of words belonging to the category S^0 was generated. Then, items from this list were combined in two different environments, and the outcomes examined for grammaticality.[14]

The first environment is the maximally simple concatenation of two S^0 words to form a new S^0 expression. Twenty such random concatenations were formed; these are listed in (28) below:

---

[14] The list was generated from the electronic corpus of naturalistic speech in Riau Indonesian (Gil and Litamahuputty, in preparation), as follows. First, a custom-produced program was run on the corpus, selecting random word tokens occurring in the text and then deleting duplicates, to produce an interim list of word types. The program was run to produce a list of 200 randomly selected words. This interim list was then inspected, and the following classes of items omitted: duplicates resulting from alternative spellings of the same word; extraneous symbols (such as the *xxx* used to transcribe unintelligible sounds); fillers and exclamations (e.g. *em* 'erm', *aduh* 'ouch'); and members of S^0/S^0 (e.g. *dari* 'from' exemplified in Section 4.5, and the participant marker *yang* discussed in Section 4.9). Resulting was a list of 157 randomly generated words that belong to S^0. From this list, the first 40 items were used to form the constructions in (28), and the next 80 items the constructions in (29). (The remaining 37 items were discarded.)

(28) a. Bagus tu
good DEM.DIST
'That's good'

b. Ada tandingan
exist compete-AUG
'There's a competition taking place'

c. Cakap habis
say finished
'He said it's all gone'

d. Raja tengok
king look
'The king is looking'

e. Tak dikawinkan
NEG PAT-marry-EP
'She wasn't married off'

f. Engkau pesta
2SG celebrate
'You're celebrating'

g. Ulang dekat
repeat close
'Do it again closer'

h. Kata lari
say run
'He said run'

i. Luar sungai
outside river
'Outside there's a river'

j. Tunggu masukkan
wait enter-EP
'Wait until he puts it in'

k. Sama bawah
together below
'It's the same down there'

l. Sebelah ngerti
one-side AG-understand
'The one just across understands'

m. Kurus teka-teki
thin riddle
'The thin guy is telling riddles'

n. Termenung jumpa
NON.AG-muse meet
'Although he was sulking, he found it'

o. Tangkapnya menjerat
catch-ASSOC AG-trap
'The one he caught was trapping'

p. Anjing giginya
dog teeth-ASSOC
'The teeth are dogs' teeth'

q. Dilemparnya ditangkap
PAT-throw-ASSOC PAT-catch
'He threw it away and then it was caught'

r. Obat dicintanya
medicine PAT-love-ASSOC
'His loving her was caused by a potion'

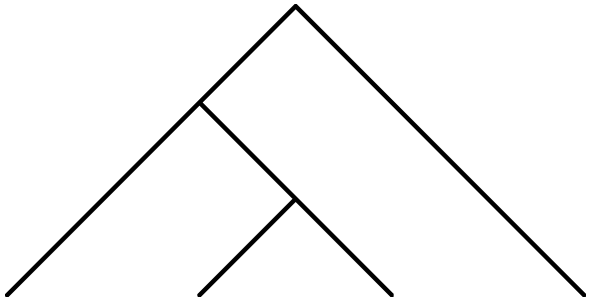
s. Nek nama
FAM-grandmother name
'Granny, it's the name'

t. Pendengaran pakai
REIF-hear-AUG use
'His hearing was with the use of it'

Inspection of the above twenty concatenations reveals that each and every one of them is completely grammatical. Of course, they differ with respect to how much sense they make, or, to put it differently, how much effort and imagination are required in order to conjure up an appropriate context in which they might be uttered; the reader can get a feel for these differences by looking at the English translations. The twenty concatenations are presented in an order, which, very loosely, reflects the degree of availability of such an appropriate context. For at least the first ten concatenations, an appropriate context comes to mind immediately, while for most of the remaining ones, such a context can be arrived at with relatively little effort. Only the last one or two concatenations are at all problematical, as reflected in their English translations, which share their strangeness. However, the problem is with the meanings, not with the forms: just as the English translations, albeit semantically anomalous, are grammatically well-formed, so the Riau Indonesian concatenations themselves, like all of the preceding ones, are completely well-formed from a morphosyntactic point of view. Thus, the little experiment represented in (28) above suggests that in Riau Indonesian, any combination whatsoever of S^0 words yields a grammatically well-formed expression.

Nevertheless, one may wonder what happens when lengthier combinations of S^0 words are formed: perhaps then certain grammatical constraints kick in, in order to rule out some potential combinations. To test this, ten strings each consisting of four randomly chosen S^0 words were formed. To make the test even more challenging, each of the ten strings was then randomly assigned one of five logically-possible binary constituencies. The resulting ten constructions are indicated in (29) below:

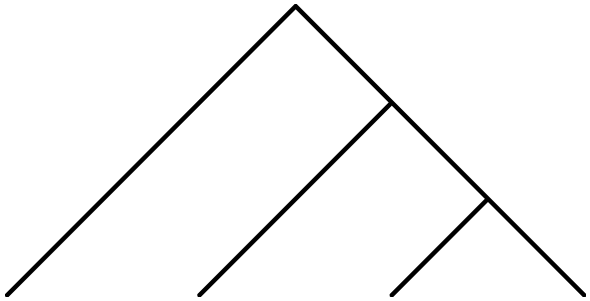
(29) a.



Bangkai	diantar	pertandingan	kejar
corpse	PAT-accompany	REIF-compete-AUG	chase

'After the corpse was brought to the competition, the chasing began'

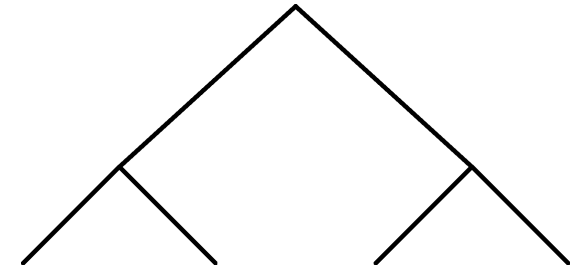
b.



Meja	bunian	dibilang	kantong
table	elf-AUG	PAT-say	bag

'As for the table, the elf was said to be in the bag'

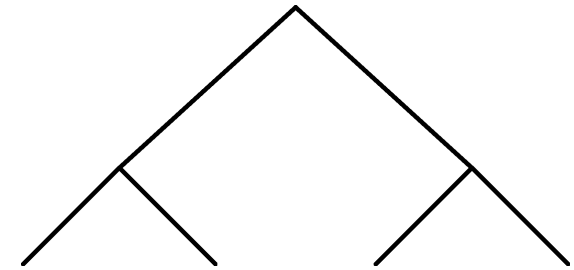
c.



Maksud	bawa	dia	sejulung
meaning	below	3	Buffon's.halfbeak

'The meaning below is that it's a Buffon's halfbeak'[15]

d.

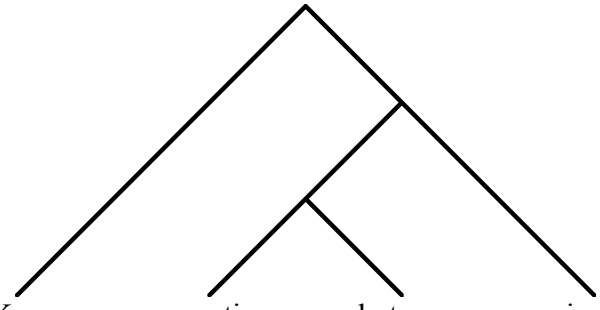


Dek	ndak	istri	boleh
FAM-younger.brother	NEG	wife	can

'Kid, no; his wife can'

[15] Buffon's halfbeak is a species of fish (Zenarchopterus buffonis).

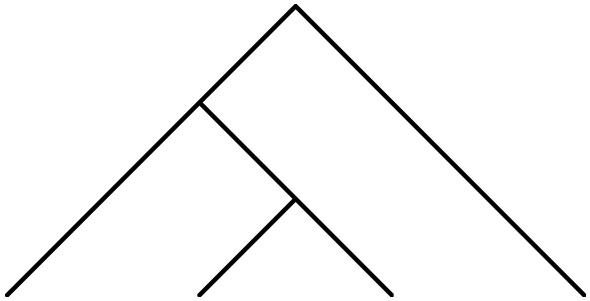
e.



Kan	nanti	batang	pinang
Q	FUT	stick	areca.palm

'You see, there'll be a tree soon and it'll be an areca palm'

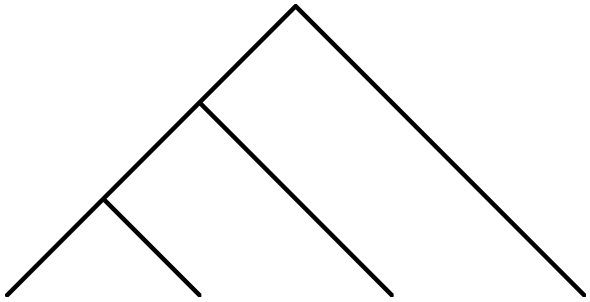
f.



Binatang	kedua	rimba	diambil
animal	ORD-two	forest	PAT-take

'He took away the animals, and secondly the forest'

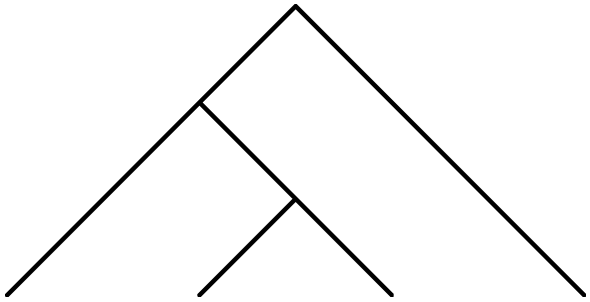
g.



Apa	jawab	dilepas	lewat
what	answer	PAT-release	pass

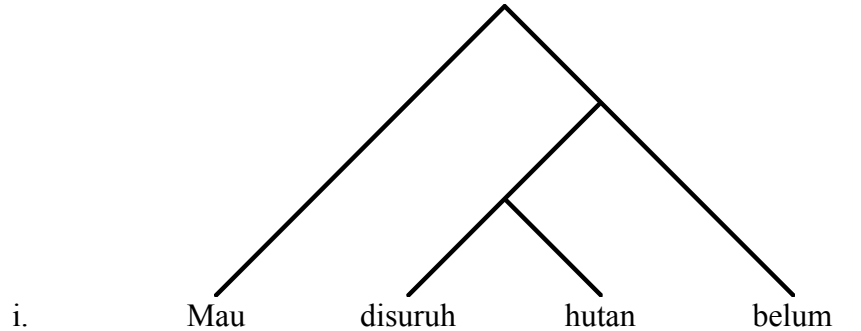
'Whatever he answers he will be released and then they will pass'

h.

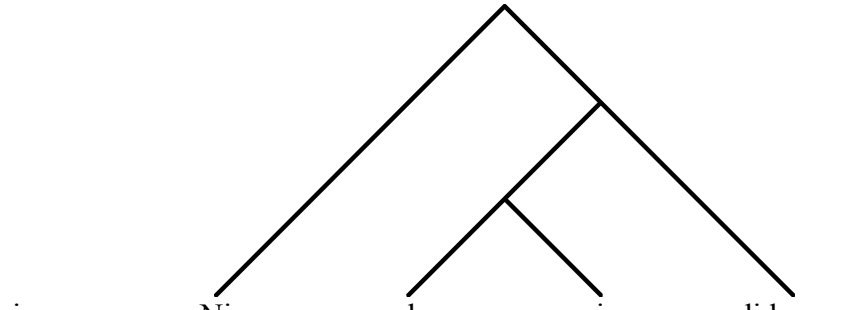


Sayuran	dapur	Inggris	anak
vegetable-AUG	kitchen	England	child

'The vegetables from the English kitchen are for the child'



'He wants them not to be ordered into the forest yet'



'You can hear him trying the long ones'

Once again, examination of the above ten concatenations reveals that each and every one of them is completely grammatical. Admittedly, in comparison with the two-word concatenations in (28), a greater effort is required in order to make sense out of them, as indeed is suggested by their translations into English. (In some cases, the same string would have been more easily associated with an appropriate context if the assigned constituent structure had been different.) Still, their strangeness is entirely semantic; in terms of their morphosyntactic structure, they are all completely well-formed.

Thus, the two little experiments represented in (28) and (29) suggest that it is indeed the case that in Riau Indonesian, any combination whatsoever of S^0 words — including, among others, thing and activity expressions — yields a grammatically well-formed expression. Accordingly, the claim that Riau Indonesian lacks a noun/verb or nominal/verbal distinction would seem to satisfy the condition of exhaustiveness.[16]

5.4 Interim summary

The claim that Riau Indonesian lacks a noun/verb or nominal/verbal distinction has thus been shown to satisfy all three criteria proposed by Evans and Osada (2005): bidirectionality, compositionality and exhaustiveness. On this basis, it may be concluded

[16] It should be noted that even if a much more extensive experiment along the same lines were to throw up the odd exception, in the form of an ungrammatical combination of expressions, this would not necessarily be damning to the description of Riau Indonesian as having a single open syntactic category S^0 and a single closed one S^0/S^0. Such exceptions could potentially be handled by removing particular words from the category S^0 and reassigning them to the closed category S^0/S^0. What would constitute refutation of the proposed description would be a body of ungrammatical combinations sufficiently large as to require the positing of a second open syntactic category.

that Riau Indonesian presents a clear case of that often proposed yet still controversial language type, namely one without a noun/verb or nominal/verbal distinction.

6 Riau Indonesian in Typological Perspective

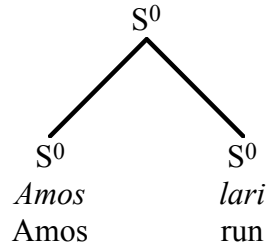
In their paper, Evans and Osada (2005) define four "ways a language could lack a noun-verb distinction", which they refer to as "omnipredicative", "precategorial", "Broschartian", and "rampart zero conversion". However, asking in what ways a language might lack a noun/verb distinction leads us right back to a perspective in which having a noun/verb distinction is taken as the default, and lacking such a distinction is considered to be the exceptional case worthy of special inquiry. But as argued in Section 1 above, this is the wrong way of looking at things. Instead, we should take the absence of distinct syntactic categories to be the default, and ask in what ways a language might override this default by having a noun/verb distinction.

This is the perspective which is adopted in the theory of syntactic categories proposed in Gil (2000c) and briefly described at end of Section 3 above. Within this categorial-grammar-based framework, syntactic categories are built up recursively, by means of category-formation operators, from a single primitive syntactic category S^0. The framework defines a derivational path for syntactic categories, in which simpler categories form the basis for the derivation of more complex ones; the derivation of a syntactic category is reflected in its name, which provides a measure of the category's complexity — longer category names being associated with more complex categories. As argued in Gil (2000c, 2005b, 2006a, 2008b), the derivation of syntactic categories is manifest in several distinct domains. *Phylogenetically*, simpler categories evolved before more complex ones; *ontogenetically*, children acquire simpler categories before more complex ones; and *typologically*, languages are more likely to have simpler categories than more complex ones. In particular, since S^0 is the simplest syntactic category, this means that phylogenetically, S^0 evolved before other syntactic categories; ontogenetically, S^0 is acquired before other syntactic categories; and typologically, S^0 is more common cross-linguistically than other syntactic categories. This latter statement can be formulated as a uni-directional implicational universal: If a language has other, more complex syntactic categories, then it also has S^0 (but not vice versa). As argued in this paper, Riau Indonesian upholds this implicational universal by providing the crucial case of a language that — among its open classes, at least — has S^0 but no other syntactic categories of greater complexity of the kind that can form the basis for a noun/verb distinction. Thus, from evolutionary, developmental and typological perspectives, the monocategorial open syntactic inventory of Riau Indonesian represents the default case.[17]

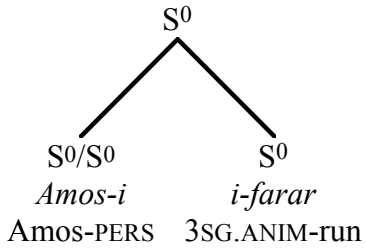
[17] Of course, this does not mean that Riau Indonesian is "primitive" in any sense. Evolutionarily, the monocategoriality of Riau Indonesian is unlikely to represent an uninterrupted inheritance from the distant pre-human past; rather, it is more plausibly viewed as resulting from the loss of syntactic-category distinctions in more recent times. On the one hand, the prevalence of apparent monocategoriality throughout the Austronesian family, including languages of quite different grammatical types, provides a prima facie case for attributing it to proto-Austronesian as spoken some 5000 years ago. On the other hand, the relative scarcity of monocategoriality across the world's languages makes it more likely that, even if proto-Austronesian were monocategorial, its own ancestors might have had more complex inventories of syntactic categories.

To see how the default case of monocategoriality may be overridden and a noun/verb distinction introduced, we may compare a basic sentence 'Amos is running' in Riau Indonesian with its counterparts in two other languages, Roon and Hebrew:[1]

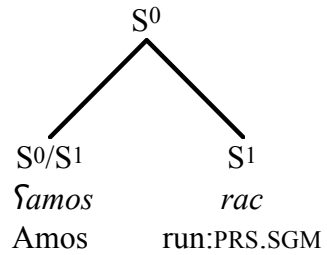
(30) a. RIAU INDONESIAN



b. ROON



c. HEBREW



In Riau Indonesian, *Amos* and *lari* are both members of S^0, and their combination thereby constitutes a coordination of S^0 expressions, and is thus itself a member of S^0. In contrast, in both Roon and Hebrew, the two words constituting the sentence belong to different syntactic categories; however, they do so in two distinct ways.

In Roon, *i-farar* and *Amos-i i-farar* both belong to the same category S^0, while *Amos-i* belongs to a different category, S^0/S^0; this captures the insight that in Roon, as in many other languages often referred to as "pronominal argument languages", full nominal expressions do not seem to be governed by the verb in the same sense as in familiar European languages, but rather stand in a looser relationship to the verb, one bearing a closer resemblance to the attributive modifier relationship characteristic of sentential adverbs. Thus, in Roon, in accordance with the definitions provided in (2) above, S^0/S^0 prototypically consists of thing expressions and is hence a nominal category, while S^0 generally contains activity expressions and is therefore a verbal category. In contrast, in Hebrew, *ʕamos*, *rac*, and *ʕamos rac* belong to three different syntactic categories, S^0/S^1, S^1, and S^0 respectively; Hebrew thus provides an appropriate exemplar for the traditional

[1] Roon is an Austronesian language of the South-Halmahera-West-New-Guinea subgroup, spoken on the island of Roon in the Cenderawasih Bay in the western part of New Guinea; the data and analysis presented here are based on my own ongoing fieldwork.

analysis of familiar European languages, represented, among others, in the well-known old-fashioned phrase-structure rewrite rule S → NP VP. In Hebrew, then, S^0/S^1 prototypically consists of thing expressions and is hence a nominal category, while S^1 generally contains activity expressions and is therefore a verbal category. Thus, although both Roon and Hebrew have a nominal/verbal distinction, their nominals and verbals actually represent different syntactic categories as defined on purely distributional grounds. What this shows, then, is that there is more than one way to have a nominal/verbal distinction.[18]

So far, the discussion has dealt exclusively with syntactic categories, defined in terms of distributional privileges. However, as argued, among others, by Hengeveld (this volume) and Hengeveld and van Lier (2009), syntactic category inventories may also correlate with morphological properties of languages. In order to better appreciate the position of Riau Indonesian within the typology of the world's languages, it is necessary to adopt a somewhat broader perspective, taking into consideration not just syntactic properties but morphological and semantic ones as well. To this end we recall the definition, first presented in Gil (2005b), of an ideal language type, that of *Isolating-Monocategorial-Associational* (or *IMA*) *Language*:

---

[18] The different syntactic structures of basic sentences in Riau Indonesian, Roon and Hebrew as represented in (30) above have far-reaching repercussions throughout the grammars of the respective languages. As in Riau Indonesian — recall the examples in (4) — in Roon and in Hebrew too, activity expressions can stand alone as complete sentences. The following sentences are obtained from those in (30) by omitting the first of the two words in each example:

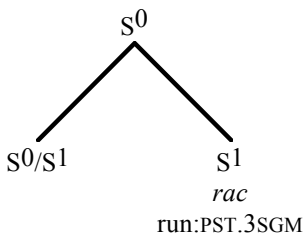
(i) a. RIAU INDONESIAN

S^0
lari
run

b. ROON

S^0
i-farar
3SG.ANIM-run

c. HEBREW



Since in Riau Indonesian and Roon, *lari* and *i-farar* are S^0 expressions, the sentences in (i/a) and (i/b) are structurally complete. In contrast, in Hebrew, *rac* is an S^1 expression, and hence, to constitute a structurally complete sentence, it must occur in construction with an S^0/S^1 position, even if that position remains empty. Thus, of the three languages under consideration, Hebrew alone merits the characterization as a "pro-drop" language; in Riau Indonesian and in Roon one-word sentences such as those in (i) above contain no empty positions and thus no "pro" to be "dropped".

(31) *Isolating-Monocategorial-Associational Language*:

a. *Morphologically Isolating*
No word-internal morphological structure;

b. *Syntactically Monocategorial*
No distinct syntactic categories;

c. *Semantically Associational*
No distinct construction-specific rules of semantic interpretation (instead, compositional semantics relies exclusively on the Association Operator).

In (31c), reference is to the Association Operator mentioned in Section 5.2 above, but generalized in order to apply polyadically, to two or more meanings in juxtaposition. For example, in the binary case, if E_1 and E_2 are two expressions with meanings M_1 and M_2 respectively, the meaning of the combination of E_1 and E_2 is A (M_1, M_2), or 'entity associated with M_1 and M_2'.

As with monocategoriality itself, IMA language is manifest in various domains. As argued in Gil (2005b, 2006a), it is present in the communicative systems of captive great apes, suggesting that the cognitive ability for IMA language was present at least as early as the common ancestors of humans and great apes some thirteen million years ago, Similarly, as argued in Gil (2005b, 2008b), it is present in the language of young infants. Clearly, however, no contemporary adult language meets any of the three conditions spelled out in (31) above: (a) every language has at least some word internal structure; (b) every language has at least some distinct syntactic categories (if only closed ones); and (c) every language has at least some construction-specific rules of semantic interpretation. The three defining properties of IMA language represent the limiting points of maximal simplicity within their respective domains: morphology, syntax and semantics. Accordingly, within each domain, languages may vary along a scale ranging from simpler to more complex. And in fact, within each of the three domains, Riau Indonesian lies near, if not at, the simple end of the scale: (a) it is strongly isolating, with no inflectional morphology, little derivational morphology, and little compounding; (b) it has but one open syntactic category S^0 and one closed syntactic category S^0/S^0; and (c) it has few construction-specific rules of semantic interpretation (such as, for example, the rule governing the interpretation of constructions containing *yang* discussed in Section 4.9). Thus, Riau Indonesian may be characterized as a *relatively IMA Language*.

The degree to which Riau Indonesian approximates pure IMA language may be appreciated by examining basic sentences such as *Amos lari* 'Amos is running' in (30a). Both words are monomorphemic; both words belong to the single open syntactic category S^0; and the semantics is entirely associational: as argued extensively elsewhere, the proper semantic representation for such a sentence is A (AMOS RUN), or 'entity associated with Amos and with running'. Example (30a) is artificially constructed; however, similar utterances exhibiting pure IMA structure occur readily in naturalistic contexts, as for example in (8a), (17a), (18a), (24b), (26a), and (27b) above. Indeed, the lion's share of the expressive power of Riau Indonesian is achieved with means that are purely IMA; what additional non-IMA levels of structure that Riau Indonesian has would seem to contribute little to its functionality as a communicative system. (This point is made in greater detail with respect to Malay/Indonesian more generally in Gil 2009).

The three properties constituting IMA language are logically independent of each other; languages can vary independently along each of the three relevant scales. These three properties thus set the stage for a three-dimensional morphological-syntactic-semantic typology of language. The outlines of such a typology are presented in Table 1:

Type	*Morphological Structure*	*Syntactic Categories*	*Compositional Semantics*	*Language*
1	Ø	Ø	Ø	Riau Indonesian
2	Ø	Ø	+	?
3	Ø	+	Ø	?
4	Ø	+	+	?
5	+	Ø	Ø	?
6	+	Ø	+	?
7	+	+	Ø	?
8	+	+	+	Hebrew

Table 1: The Isolating-Monocategorial-Associational (IMA) Typology

To make the typology manageable, location along each of the three scales of complexity is reduced to a binary distinction, with "Ø" representing a position near the simple end of the scale, corresponding to isolating, monocategorial and associational structure respectively, and "+" representing a position significantly higher with respect to complexity. Thus, Type 1, representing relative IMA language, is exemplified with Riau Indonesian, while Type 8, representing languages that are significantly more complex in all three domains, can be exemplified by Hebrew and very many other familiar languages. The use of the "Ø" and "+" symbols embodies the perspective whereby, for each scale, the default value is the simple one, here indicated with "Ø"; in accordance with this perspective, relative IMA language, as exemplified by Riau Indonesian, constitutes the default language type.

The IMA Typology in Table 1 makes it possible to ask whether any of the remaining six logical possibilities are actually attested. The answer is almost certainly yes, though the task of filling in the empty cells poses a serious analytical challenge. Estimating how much morphology a language has is relatively more straightforward, though even here there is a grey border area consisting of clitics and possible compounds whose status as independent words can be difficult to adjudicate. However, determining the inventory of syntactic categories of a language is notoriously difficult, as witnessed by the numerous controversies in the literature, which can be attributed to a combination of two factors: the need for any analyses to be based on detailed and painstaking descriptions of the relevant grammatical patterns, and the inevitable dependence of such analyses on theoretical assumptions and persuasions which may differ from one practitioner to another. And assessing the complexity of associational semantics in a language is even more challenging. In work in progress, some preliminary results of which are reported in Gil (2007, 2008a), a cross-linguistic experiment is conducted, designed to measure the degree to which the compositional semantics of different languages relies on the association operator, to the exclusion of other construction-specific rules of semantic interpretation. However, even this experiment provides a measure of just one aspect of associational semantics, pertaining to thematic roles at the clausal level; other domains of compositional semantics are not examined.

Nevertheless, one may speculate as to how some of the intermediate Types 2 - 7 of the IMA Typology might be exemplified. Most of the other languages that have been argued to be monocategorial differ from Riau Indonesian in that they are not isolating. An example of such a language is Tagalog, argued to be lacking a noun/verb distinction in Gil (1993b,c, 1995). Not only does it have rich morphological structure, but it would also

appear to be considerably less associational than Riau Indonesian. For example, a basic Tagalog sentence such as *Tumatakbo si Amos* 'Amos is running' is more semantically specific than its Riau Indonesian counterpart in (30a) in at least two important respects. First, it can only be understood predicatively, unlike its Riau Indonesian counterpart, which can also be interpreted attributively, to mean 'The Amos who is running'. Secondly, the voice infix *-um-* in conjunction with the personal marker *si* ensures that 'Amos' can only be understood as the actor of 'run', unlike the corresponding sentence in Riau Indonesian, where 'Amos' can bear any thematic role that makes sense in the given context, such as cause or goal. Thus, in Tagalog, associational semantics is overridden by additional rules of compositional semantics relating specific morphosyntactic features to notions such as predication and thematic roles. Accordingly, Tagalog provides a likely candidate for a Type 6 language: monocategorial, but neither isolating nor associational.[19] In conjunction, Tagalog and Riau Indonesian show how monocategoriality and the absence of a noun/verb distinction are independent — not just logically but also empirically — of morphological complexity and of complexity in the domain of compositional semantics.

Turning now to the three remaining types of languages with more complex inventories of syntactic categories, it is clear that these may differ with respect to morphological and semantic complexity. Type 3, isolating and with associational structure, may perhaps be instantiated by Ju|'hoan; Type 4, isolating but with non-associational semantics, might possibly be exemplified by Papiamentu; while Type 7, non-isolating but with associational semantics, is potentially represented by Meyah.[20] To the extent that such preliminary tentative characterizations are borne out by further investigations, the likelihood increases that most or all of the eight types defined by the IMA Typology may indeed be attested.

The IMA typology suggests that, as a relative IMA language, Riau Indonesian occupies a privileged default position in the typological space defined by patterns of cross-linguistic variation. One may wonder how and why it comes about that a certain language is endowed with such special status. A possible response is that the apparent distinctiveness of Riau Indonesian is just an artifact of the criteria with respect to which it was examined: pick a different set of criteria, and some other language or languages will seem to represent the default type. To a limited degree this may be the case. Nevertheless, the criteria on which the IMA Typology is based are so central to grammatical organization that is hard to believe that they do not reflect a very basic property of languages.

Assuming, then, that there is indeed something real about the default nature of Riau Indonesian, one may then ask what kinds of diachronic processes led it to where it is now. A possible answer, proposed by McWhorter (2001, 2005, 2006, 2008), is that Riau Indonesian has undergone massive structural simplification due to language contact and pervasive second-language acquisition. In earlier writings, McWhorter suggests that Riau Indonesian is a creole language; however, as pointed out in Gil (2001a), there is no independent evidence for such a claim. In more recent writings, however, McWhorter argues that Riau Indonesian, and in fact Malay/Indonesian in general, exemplifies what he calls a *Non-Hybridized Conventionalized Second Language* (NCLS), a language with a

---

[19] As for the remaining two types of monocategorial languages, Types 2 and 5, at present I am familiar with no plausible candidates to represent these types. Further investigations may or may not reveal the existence of such languages.

[20] The characterizations of these languages as having, at the very least, a noun/verb distinction, are taken from published descriptions which make explicit and seemingly unavoidable reference to these categories: Güldemann (2000) for Ju|'hoan, Dijkhoff (1993) for Papiamentu, and Gravelle (2004) for Meyah. The varying attributions of associationality to these languages reflect the results of my own fieldwork and the association experiment.

large population of speakers, which, due to rampant second-language acquisition, tends to be simpler than its non-NCLS relatives; some other examples of NCLSs include Hindi/Urdu, Arabic and English. While it may in fact be the case that some of the simplicity of Malay/Indonesian is due to its status as an NCLS, this cannot be the whole story. In particular, it does not explain why close but non-NCLS relatives of Malay/Indonesian such as Minangkabau are comparably simple, or why other NCLSs, such as Hindi/Urdu, Arabic and English, are so much more complex.

Given that the three criteria underlying the IMA typology are logically and empirically independent of each other, it is probably more appropriate to view Riau Indonesian as what you would expect to get when three rolls of the relevant typological dice just happen to produce three "Ø"s in succession. Similarly, in geographical perspective, one may consider Riau Indonesian to lie at the fortuitous intersection of three partly-overlapping linguistic regions: that of isolating languages, centered in mainland Southeast Asia but extending downwards through Sumatra and Java into the islands of Nusa Tenggara; that of monocategorial languages, comprising much of Austronesian and possibly other languages of the northern Pacific rim; and that of associational languages, apparently encompassing much of the Indonesian archipelago. Looking at things in this way, Riau Indonesian represents nothing more exceptional than one of a number of possible combinations of values of independently varying linguistic features.

How many other languages are there like Riau Indonesian? Work in progress suggests that amongst colloquial Malay/Indonesian dialects, Riau Indonesian is anything but exceptional. In spite of the numerous salient differences between Malay/Indonesian varieties, comparable perhaps to those between dialects of Arabic, most varieties of Malay/Indonesian share the same typological ground plans. My impression, based on familiarity with some half a dozen Malay/Indonesian varieties from across the archipelago, is that most or all of what has been said in this paper about Riau Indonesian holds, with just minor modifications, for most or all other varieties of colloquial Malay/Indonesian. In addition, it is quite likely that, in the region where Malay/Indonesian is spoken, there are other languages that share the same broad typological characteristics. Outside the insular Southeast region, however, I am at present not familiar with any clear cut examples of relatively IMA languages. This could be because there are none, but I cannot help but suspect that they do exist — it's just that we are not looking for them the right way, not approaching them in the right frame of mind.

7 Conclusion

When working on a new and unfamiliar language and the facts seem to point in a certain direction, we don't generally say, "Hmm, maybe it doesn't have a dual". Rather, if the facts appear to be otherwise, we cry out "Hey, it has a dual"! By the same token, when working on a new language and it turns out that words like 'eat' and 'chicken' exhibit the same grammatical behaviour, we should not scratch our heads in bewilderment, do everything we can to try and find differences, and only after failing in this quest, reluctantly conclude "Oh well, maybe it doesn't have a noun/verb distinction." Instead, our approach to the language should be such that if we *do* encounter systematic differences in grammatical behaviour between words denoting things and words denoting activities, we should then exclaim "Aha, it has a noun/verb distinction!" As reported on in detail in this paper, Riau Indonesian provides no cause for such an exclamation.

Acknowledgements

This paper owes its existence to the many speakers of Riau Indonesian whose speech found its way into the naturalistic data cited in this paper: Aidil, Anton, Arief, Arip Kamil, Bismar, Botak, Elly Yanto, Erimulia, Jamil, Kairil, Kartini, Methandra Saputra, Pai, Riki, Rudi Chandra, Septianbudiwibowo, Syafi'i, Syarifudin, Yendy Ardiansyah, Zainudin, and others. It has also benefited from work by Santi Kurniati and Yessy Prima Putri, who transcribed and annotated the the Max Planck Institute Padang Field Station Corpus of Riau Indonesan (Gil and Litamahuputty in preparation), and by Brad Taylor, who wrote the program generating random word lists from the electronic corpus. Various bits and pieces of this paper have been presented at numerous talks throughout the years, at which many of the participants — too many to acknowledge in person — made valuable comments and suggestions.

References

Aliana, S. A., Nursato, S., Arifin, S. S., Soetopo, S., and Waif, M. (1987). *Morfologi dan Sintaksis Bahasa Melayu Palembang*, Jakarta: Pusat Pembinaan dan Pengembangan Bahasa, Departemen Pendidikan dan Kebudayaan.

Beck, D. (2002). *The Typology of Parts of Speech Systems, The Markedness of Adjectives*. New York: Routledge.

Broschart, J. (1997). 'Why Tongan does it differently: categorial distinctions in a language without nouns and verbs', *Linguistic Typology* 1: 123-165.

Carnie, A. (1997). 'Two types of non-verbal predication in Modern Irish', *Canadian Journal of Linguistics* 42: 57-73.

Croft, W. (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.

--- (2003). *Typology and Universals*, 2nd Edition. Cambridge: Cambridge University Press.

Dijkhoff, M. B. (1993). *Papiamentu Word Formation, A Case Study of Complex Nouns and their Relation to Phrases and Clauses*. PhD dissertation, University of Amsterdam.

Dixon, R.M.W. and Aikhenvald, A. Y. (eds.) (2006). *Adjective Classes: A Cross-Linguistic Typology*. Oxford: Oxford University Press.

Dryer, M. S (1997). 'Are grammatical relations universal?', in J. Bybee, J. Haiman and S.A. Thompson (eds.), *Essays on Language Function and Language Type.* Amsterdam: Benjamins, 115-143.

Ewing, M. C. (2005). 'Colloquial Indonesian', in A. Adelaar and N.P. Himmelmann (eds.), *The Austronesian Languages of Asia and Madagascar*. London and New York: Routledge, 227-258.

Evans, N. (2000). 'Kinship verbs', in P. M. Vogel and B. Comrie (eds.), *Approaches to the Typology of Word Classes* (Empirical Approaches to Language Typology 23). Berlin and New York: Mouton de Gruyter, 103-172.

Evans, N. and Osada, T. (2005). 'Mundari: the myth of a language without word classes', *Linguistic Typology* 9: 351-390.

François, A. (2010). 'Pragmatic demotion and clause dependency: On two atypical subordinating strategies in Lo-Toga and Hiw (Torres, Vanuatu)', in I. Brill (ed.), *Clause Hierarchy and Clause Linking: The Syntax and Pragmatics Interface* (Studies in Language Companion Series 121). Amsterdam: Benjamins.

Gil, D. (1993a). Nominal and verbal quantification', *Sprachtypologie und Universalienforschung* 46: 275-317.

--- (1993b). Syntactic categories in Tagalog", in S. Luksaneeyanawin (ed.), *Pan-Asiatic Linguistics, Proceedings of the Third International Symposium on Language and Linguistics*, Chulalongkorn University, Bangkok, January 8-10, 1992, Volume 3, 1136-1150.
--- (1993c). 'Tagalog semantics', in J.S. Guenter, B.A. Kaiser and C.C. Zoll (eds.), *Proceedings of the Nineteenth Annual Meeting of the Berkeley Linguistics Society, February 12-15, 1993*, Berkeley Linguistics Society, Berkeley, 390-403.
--- (1994). 'The structure of Riau Indonesian', *Nordic Journal of Linguistics* 17: 179-200.
--- (1995). 'Parts of Speech in Tagalog', in M. Alves (ed)., *Papers from the Third Annual Meeting of the Southeast Asian Linguistics Society*, Arizona State University, Tempe, 67-90.
--- (1999). 'Riau Indonesian as a pivotless language', in E.V. Raxilina and Y.G. Testelec (eds.), *Tipologija i Teorija Jazyka, Ot Opisanija k Objasneniju, K 60-Letiju Aleksandra Evgen'evicha Kibrika (Typology and Linguistic Theory, From Description to Explanation, For the 60th Birthday of Aleksandr E. Kibrik)*, Jazyki Russkoj Kul'tury, Moscow, 187-211.
--- (2000a). 'Island-like phenomena in Riau Indonesian', Paper presented at the *Seventh Annual Meeting of the Austronesian Formal Linguistics Association*, Amsterdam, The Netherlands, 12 May 2000.
--- (2000b). 'Riau Indonesian: A VO Language with Internally-Headed Relative Clauses1, *Snippets* 1.
--- (2000c). 'Syntactic categories, cross-linguistic variation and universal grammar', in P. M. Vogel and B. Comrie (eds.), *Approaches to the Typology of Word Classes* (Empirical Approaches to Language Typology 23). Berlin and New York: Mouton de Gruyter, 173-216.
--- (2001a). 'Creoles, complexity and Riau Indonesian', *Linguistic Typology* 5: 325-371.
--- (2001b). 'Escaping Eurocentrism: Fieldwork as a Process of Unlearning', in P. Newman and M. Ratliff (eds.), *Linguistic Fieldwork*. Cambridge: Cambridge University Press, 102-132.
--- (2001c). 'Reflexive anaphor or conjunctive operator? Riau Indonesian *sendiri*', in P. Cole, C.-T. J. Huang and G. Hermon (eds.), *Long Distance Reflexives* (Syntax and Semantics 33). New York: Academic Press, 83-117.
--- (2002a). 'Ludlings in Malayic languages: An introduction', in B. K. Purwo (ed.), *PELBBA 15, Pertemuan Linguistik Pusat Kajian Bahasa dan Budaya Atma Jaya: Kelima Belas*, Jakarta: Unika Atma Jaya, 125-180.
--- (2002b. 'The prefixes *di-* and *N-* in Malay / Indonesian dialects', in F. Wouk and M. Ross (eds.), *The History and Typology of Western Austronesian Voice Systems*. Canberra: The Australian National University, 241-283.
--- (2003a). 'English goes Asian; number and (in)definiteness in the Singlish noun-phrase', in F. Plank (ed.), *Noun Phrase Structure in the Languages of Europe* (Empirical Approaches to Language Typology, Eurotyp 20-7). Berlin and New York: Mouton de Gruyter, 467-514.
--- (2003b). 'Intonation does not differentiate thematic roles in Riau Indonesian', in A. Riehl and T. Savella (eds.), *Proceedings of the Ninth Annual Meeting of the Austronesian Formal Linguistics Association FLA9), Cornell Working Papers in Linguistics* 19: 64-78.
--- (2004a). 'Learning about language from your handphone; *dan, and* and *&* in SMSs from the Siak River Basin', in K. E. Sukatmo (ed.), *Kolita 2, Konferensi Linguistik Tahunan Atma Jaya,* Jakarta: Pusat Kajian Bahasa dan Budaya, Unika Atma Jaya, 57-61.

--- (2004b). 'Riau Indonesian *sama*, explorations in macrofunctionality', in M. Haspelmath (ed.), *Coordinating Constructions* (Typological Studies in Language 58). Amsterdam: Benjamins, 371-424.
--- (2005a). 'From repetition to reduplication in Riau Indonesian', in B. Hurch (ed.) *Studies on Reduplication* (Empirical Approaches to Language Typology 28). Berlin – New York: Mouton de Gruyter, 31-64.
--- (2005b). 'Isolating-monocategorial-associational language', in H. Cohen and C. Lefebvre (eds.), *Categorization in Cognitive Science*. Oxford: Elsevier, 347-379.
--- (2005c). 'Word order without syntactic categories: how Riau Indonesian does it', in A. Carnie, H. Harley and S.A. Dooley (eds.), *Verb First: On the Syntax of Verb-Initial Languages*. Amsterdam: Benjamins, 243-263.
--- (2006a). 'Early human language was isolating-monocategorial-associational', in A. Cangelosi, A.D.M Smith and K. Smith (eds.), *The Evolution of Language, Proceedings of the 6th International Conference (EVOLANG6)*. Singapore: World Scientific, 91-98.
--- (2006b). 'Intonation and thematic roles in Riau Indonesian', in C.M. Lee, M. Gordon, and D. Büring (eds.), *Topic and Focus, Cross-Linguistic Perspectives on Meaning and Intonation* (Studies in Linguistics and Philosophy 82). Dordrecht: Springer, 41-68.
--- (2007). 'Creoles, complexity and associational semantics', in U. Ansaldo and S.J. Matthews (eds.), *Deconstructing Creole: New Horizons in Language Creation*. Amsterdam: Benjamins, 67-108.
--- (2008a). 'How complex are isolating languages?' in F. Karlsson, M. Miestamo and K. Sinnemäki (eds.), *Language Complexity: Typology, Contact, Change*. Amsterdam: Benjamins, 109-131.
--- (2008b). 'The Acquisition of Syntactic Categories in Jakarta Indonesian', *Studies in Language* 32: 637-669.
--- (2009). 'How much grammar does it take to sail a boat?', in G. Sampson, D. Gil and P. Trudgill (eds.), *Language Complexity as an Evolving Variable*. Oxford: Oxford University Press,19-33.
--- (to appear). 'What Is Riau Indonesian?'.
Gil, D. and Litamahuputty, B. (in preparation). *The MPI Riau Corpus,* A joint project of the Department of Linguistics, Max Planck Institute for Evolutionary Anthropology, the Center for Language and Culture Studies, Atma Jaya Catholic University, and Universitas Bung Hatta.
Gravelle, G. G. (2004). *Meyah: An East Bird's Head Language of Papua, Indonesia*, PhD Dissertation, Free University of Amsterdam.
Güldemann, T. (2000). 'Noun categorization systems in Non-Khoe lineages of Khoisan, *AAP 63*. Köln: Institut für Afrikanistik, Universität zu Köln, 5-33.
Haspelmath, M. (2007). 'Pre-established categories don't exist - consequences for language description and typology', *Linguistic Typology* 11: 119-132.
Hengeveld, K. (1992). *Non-verbal predication: theory, typology, diachrony* (Functional Grammar Series 15). Berlin and New York: Mouton de Gruyter.
--- (this volume). 'Parts-of-speech system as a basic typological determinant'.
Hengeveld, K. and Lier, E. H. van (2010). 'An implicational map of parts-of-speech', *Linguistic* Discovery 8-1: 129-156.
Jelinek, E. and Demers, R. A. (1994). 'Predicates and pronominal arguments in Straits Salish, *Language* 70: 697-736.
Macdonald, R. and Darjowidjojo, S. (1967). *A Student's Reference Grammar of Modern Formal Indonesian*. Washington: Georgetown University Press.
McWhorter, J. (2001). 'What people ask David Gil and why', *Linguistic Typology* 5: 388-412.

McWhorter, J. (2005). *Defining Creole*. Oxford and New York: Oxford University Press.
--- (2006). *Language Interrupted: Signs of Non-Native Acquisition in Standard language Grammars*. New York: Oxford University Press.
--- (2008.) 'Why does a language undress? Strange cases in Indonesia', in F. Karlsson, M. Miestamo and K. Sinnemäki (eds.), *Language Complexity: Typology, Contact, Change*. Amsterdam: Benjamins, 167-190.
Minde, D. van (1997). *Malayu Ambong*. Leiden: Research School CNWS, School of Asian, African and Amerindian Studies.
Mintz, M. W. (1994). *A Student's Grammar of Malay and Indonesian*. Singapore: EPB Publishers.
Mosel, U. and Hovdhaugen, E. (1992). *Samoan reference grammar* (Instituttet for sammenlignende kulturforskning, Series B: Skrifter, LXXXV). Oslo: Scandinavian University Press.
Nik Safiah Karim, Farid M. Onn, Hashim Hj. Musa, and Abdul Hamid Mahmood (1996). *Tatabahasa Dewan, Edisi Baharu*, Kuala Lumpur: Dewan Bahasa dan Pustaka, Kementrian Pendidikan Malaysia.
Nordhoff, S. (this volume). 'Jack-of-All-Trades: The Sri Lanka Malay flexible adjective'.
Shkarban, L. I. (1992). Syntactic aspect of part-of-speech typology', in S. Luksaneeyanawin (ed.), *Pan-Asiatic Linguistics, Proceedings of the Third International Symposium on Language and Linguistics* (Vol. 1) Bangkok: Chulalongkorn University, 261-275.
--- (1995). *Grammatičeski Stroj Tagal'skogo Jazyka*. Moscow: Vostočnaja Literatura.
Sasse, H.-J. (1993). 'Syntactic categories and subcategories', in J. Jacobs et al. (eds.), *Syntax* (Handbücher zur Sprach- und Kommunikationswissenschaft 9.1). Berlin and New York: Mouton de Gruyter, 646-686.
Simin, A. M. (1988). *Discourse-Syntax of "Yang" in Malay (Bahasa Malaysia)*. Kuala Lumpur: Dewan Bahasa dan Pustaka, Kementerian Pendidikan Malaysia.
Sneddon, J. N. (1996). *Indonesian Reference Grammar*. Crows Nest NSW: Allen & Unwin.
--- (2006). *Colloquial Jakartan Indonesian*. Canberra: The Australian National University.
Stassen, L. (1997). *Intransitive Predication*. Oxford: Clarendon.
Stoel, R. (2005). *Focus in Manado Malay*. Leiden: Research School CNWS, School of Asian, African and Amerindian Studies.
Swadesh, M. (1939). 'Nootka internal syntax', *International Journal of American Linguistics* 9: 77-102.
Tchekhoff, C (1984). 'Une langue sans opposition verbo-nominale: le Tongien', *Modèles linguistiques* 6: 125-132.
Thompson, L. C. (1965) *A Vietnamese Grammar*. Seattle: University of Washington Press.
Tjung, Y. (2006). *The Formation of Relative Clauses in Jakarta Indonesian: A Subject-Object Asymmetry*. PhD dissertation, University of Delaware, Newark.
Vogel, P. and Comrie, B. (eds). (2000). *Approaches to the Typology of Word Classes*, (Empirical Approaches to Language Typology 23). Berlin and New York: Mouton de Gruyter.

Parts of speech in Kharia: a formal account*

John Peterson

1 Introduction

In previous studies (e.g. Peterson 2005; 2006; 2007; 2011; Peterson and Maas 2009) I argue that Kharia, a Munda language spoken in eastern-central India, does not possess lexical parts of speech such as noun, adjective and verb. Rather, I argue that the lexicon can simply be divided into two classes: contentive and functional morphemes. While functional morphemes are limited to their respective grammatical functions, contentive morphemes can appear in referential, attributive and predicative function without any derivational morphology, light verb, etc.

The present study does not attempt to show once again that Kharia does not possess lexical classes such as noun, verb and adjective. Rather, its primary aim is to demonstrate that an analysis of Kharia which does not assume the presence of these lexical classes – one which I believe is not only descriptively adequate but also the most economical analysis – can also be formalized relatively easily within a monostratal theoretical framework.

As detailed argumentation against assuming lexical parts of speech in Kharia has appeared elsewhere, section 2.1 presents a rather brief discussion of this issue for Kharia, while sections 2.2-2.4 deal with the status and structure of predicates and their complements (≈ arguments and adjuncts) in somewhat more detail. Section 2.5 summarizes this discussion. Section 3 then presents a formalization of the analysis discussed in section 2: After introducing the framework used in the present study (3.1) and the relevance of the Lexical Integrity Principle for some theoretical frameworks (3.2), there follows a formal account of the structure of "Case-syntagmas", roughly corresponding to NPs (3.3), and "TAM/Person-syntagmas", roughly corresponding to predicates (3.4). Finally, section 4 provides a summary and outlook.

2 Kharia – a brief overview

2.1 General introduction

Kharia is a member of the southern branch of the Munda family, which forms the western branch of the Austro-Asiatic phylum. It is predominantly predicate-final, and grammatical categories such as person/number/honorific status, tam and case are marked either by enclitics or postpositions. In addition, Kharia, like many other Munda languages, has basic voice, i.e., there is an active/middle distinction in most tam categories.

There is little if any evidence in Kharia for the presence of lexical categories such as noun, adjective and verb.[1] Instead, the lexicon can be divided into contentive morphemes,

* The present study is based on the results of approximately eight months of field work conducted during five trips to Jharkhand, India. I would like to express my gratitude to the German Research Council (*Deutsche Forschungsgemeinschaft*) for two generous grants which made two of these trips possible (PE 872/1-1, 2). Furthermore, many thanks to Utz Maas, Eva van Lier, Jan Rijkhoff and two anonymous referees for comments on an earlier version of this study, although I alone am of course responsible for any errors.

[1] More precisely, I have been able to make out two "dialects" of Kharia which, although they do not seem to differ significantly with respect to phonology, do show quite distinct characteristics with respect to parts of speech. I refer to these in Peterson (2011: 76-77) as the Simdega and Gumla dialects. The Gumla dialect does seem to possess some nouns and verbs, although the rest of the lexicon is underspecified in this respect.

which can be used in predicative, referential and attributive functions, and functional morphemes, which are specialized for particular grammatical functions and as a general rule are enclitic. The following presents two simple examples demonstrating the flexibility of contentive morphemes:

(1) a. *lebu ɖel=ki.*
man come=m.pst
'The / a man came.'

b. *bhagwan lebu=ki ro ɖel=ki.*
God man=m.pst and come=m.pst
'God became man [= Jesus] and came [to earth].'
[(1b) adapted from Malhotra 1982:136]

(2) (in a play about me and you, in which both of us will be taking part):
"naʈak=te iɲ=ga ho=kaɽ=na=iɲ ro am=ga iɲ=na=m."
play=OBL 1sg=FOC that-SG.HUM=M.IRR=1SG and 2SG=FOC 1SG=M.IRR=2SG

"umbo?. am=na um=iɲ pal=e. ɖirekʈar seŋ=ga? iɲ=te
no 2SG=INF NEG=1SG be.able=A.IRR director early=FOC 1SG=OBL
ho=kaɽ=o?. am=ga am=na=m
that-SG.HUM=A.PST." 2SG=FOC 2SG=M.IRR=2SG
'"In the play I will be him and you will be me."
"No. I can't be you. The director already made me him. You will be you."'

In order to give the reader some idea of the degree of flexibility of contentive morphemes in Kharia, the following presents a few examples which show that such flexibility is not restricted to any particular semantic class(es) but is entirely productive, given a proper context. For further discussion and examples, cf. Peterson (2011, chapter 4; note: the following putative classes are purely intuitive and have no further theoretical implications):

(3) Interrogatives: *i* 'what; which; do what?'
(4) Indefinites: *jahã* 'something; some (attributive); do something'
(5) Quantifiers: *moɲ* 'one (referential/attributive); become one'
(6) Properties: *rusuŋ* 'red one; red (attributive); become red', *maha* 'big one; big; grow, become big'
(7) Proper names: *a?ghrom* 'Aghrom (name of a town) (referential/attributive); come to be called "Aghrom" (middle voice), name [something] "Aghrom" (active voice)'
(8) Status and Role: *ayo* 'mother; become a mother (middle voice), accept someone as a mother (active voice)'
(9) Deictics and proforms: *iɖa?* 'yesterday; become yesterday (middle), turn (e.g. today) into yesterday (active, e.g., with God as subject)'

Simdega Kharia, on the other hand shows no signs of lexical specialization whatsoever with respect to its contentive morphemes. The present study deals exclusively with Simdega Kharia.

The fact that Simdega Kharia, which has been in direct contact with Mundari (North Munda) for centuries, does not possess lexical categories such as noun, verb and adjective, whereas Gumla Kharia – like other South Munda languages – does, strongly suggests that this trait of Simdega Kharia has arisen through contact with speakers of Mundari, for which it is generally assumed that all lexical roots are flexible (cf. e.g. Hengeveld and Rijkhoff 2005; Hoffmann 1903, 1905/1909; Peterson 2005; for a very different analysis, however, see Evans and Osada 2005).

(10) Physical objects and animate entities: *kaɖoŋ* 'fish; become a fish (middle), turn into a fish (active)'
(11) Locative: *tobluŋ* 'top, rise (middle), raise (active)'
(12) Activities: *silo?* 'plowing (n.); plowed; plow'

In addition to this lexical flexibility, what appear to be entire NPs can also function predicatively to denote an event:

(13) *ho rocho?b=te col=ki=ɲ*
that side=OBL go=M.PST=1SG
'I went to that side'

(14) *ho rocho?b=ki=ɲ*
that side=M.PST=1SG
'I went to that side' (lit.: "I that-side-d")

At first glance, example (13) would appear to lend itself quite well to an analysis of Kharia as a language with a noun/verb distinction. However, as the construction in (14) shows, such an analysis is problematic: To begin with, *rocho?b* is not an especially "verby" morpheme, which for many might in itself be reason enough to reject such an analysis. But even if we were to treat *rocho?b* 'side' as a verb, thereby preserving our noun/verb distinction, we would then have a verb modified by the demonstrative *ho* 'that'.

Although a demonstrative modifying a verb may not in itself appear especially troubling to some, this analysis would still overlook another, more important aspect of Kharia grammar: If we insist on speaking of "nouns" and "verbs", the presumed "verb" in (14) is not a "noun (*rocho?b*) used as a verb" but rather an entire "NP", i.e., *ho rocho?b*. This is especially troubling for those who see zero-derivation at work here. Although one could argue for this in the case of simple lexemes (or rather, stems), (14) is of an entirely different nature: If we were to assume zero-derivation here, we would not be zero-deriving a verb from a nominal stem but rather from a full-fledged NP – that is, from a SYNTACTIC unit.

As the following examples show, the construction in (14), with an apparent NP as the semantic base of the predicate, is entirely productive and can also contain both quantifiers and genitive attributes, in addition to demonstratives.

(15) *ubar rocho?b=ki=ɲ*
two side=M.PST=1SG
'I moved to both sides (i.e., this way and then that).'

(16) *ayo=ya?=yo?*
mother=GEN=A.PST
'he or she made [it] mother's' (lit."he or she mother's-ed [it]")

(17) a. *iɲ ho=kaʈ=te iɲ=a?=yo?j.*
1SG that=SG.HUM=OBL 1SG=GEN=A.PST.1SG
'You adopted him/her.'

b. *am ho=kaʈ=te* *am=a?=yoˀb.*
2SG that=SG.HUM=OBL 2SG=GEN=A.PST.2SG
'I adopted him/her (i.e., I made him/her mine).'

(18) *bharat=ya? lebu=ki bides=a? lebu=ki=ya? rupraŋ=ki=may.*
India=GEN person=PL abroad=GEN person=PL=GEN appearance=M.PST=3PL
'The Indians took on the appearance of foreigners (e.g. by living abroad so long).'

(19) *o?=ya? teloŋ=ki.* *o?=ya? teloŋ=o?=ki.*
house=GEN roof=M.PST house=GEN roof=A.PST=PL
'The house's roof was thatched.' 'They thatched the house's roof.'

Enclitic grammatical markers[2]
Structures such as those in (14)-(19) are possible not only because of the flexibility of contentive morphemes but also because grammatical markers in Kharia in general are enclitic and attach to the last element of a clausal constituent, regardless of its status. Consider the following examples:

(20) a. *munu?siŋ rochob=a? lebu=ki=ko*
east side=GEN person=PL=CNTR
'the people of the east; the easterners'

b. *munu?siŋ rochob=a?=ki=ko*
east side=GEN=PL=CNTR
'the easterners; the [ones] of the east'

When it is clear from context that e.g. the speaker in (20)a is referring to the people of the east, (or alternatively, if he or she considers this information unimportant), then the presumed head lexeme *lebu* 'person' can be ommitted, resulting in the attested example in (20)b, where the plural and contrastive enclitics attach to what appears to be a genitive attribute. As examples (14)-(20) as well as the following show, this is true of virtually all types of grammatical marking in Kharia,[3] as these enclitic markers (**bold print**) all attach to the last element of a complex syntactic unit (underlined in the following examples).

- Case and number markers

(21) *la? u sembho ro ɖakay rani**=kiyar=a?*** *nãw jan*
then this Sembho and Dakay Queen=DU=GEN nine class
beˀʈ=ɖom=kiyar *aw=ki=kiyar.*
son=3POSS=HON QUAL=M.PST=HON
'And this Sembho and queen Dakay had nine sons (hon) (= 'This Sembho and Queen Dakay's nine sons were).'

[2] The present discussion of clitics must remain rather general for reasons of space. For further detail, see Peterson (2011, chapters 2 and 3) where it is shown that these elements in Kharia are both what Anderson (2005) refers to as "phonological clitics" as well as "morphosyntactic clitics".
[3] With the exception of the causative marker and a few derivational suffixes (cf. Peterson 2011: 68-69), and postpositions and the optative marker *guɖu?* (see following main text), which are both phonological words as well as syntactic atoms.

(22) *iŋ=a?=**te*** (cf. *iŋ=a? bo?=**te***)
1SG=GEN=OBL 1SG=GEN place=OBL
'at my place'

- Markers for inalienable possession

(23) *ayo aba ro boker kulam=**ɖom**=ki=ya?*
mother father and brother.in.law brother=3POSS=PL=GEN
'their mother, father, brother-in-law and brother's'

- tam and Subject-Markers (cf. also (48)-(50))

(24) a. *kayom=ta=**ŋ*** b. *um=**iŋ** kayom=ta*
speak=M.PRS=1SG NEG=1SG speak=M.PRS
'I speak' 'I do not speak'

This issue will again be of relevance when discussing the status of "words" in Kharia in section 3.2, as enclitics are "syntactic atoms" (DiSciullo and Williams 1987), "grammatical words" (Dixon and Aikhenvald 2002) or "phrasal affixes" (Sadock 1991), although not phonological words.

2.2 TAM/Person- and Case-syntagmas

In order to distinguish clearly between the functional categories of "predicate" and "referent" on the one hand and the two purely structural categories which are the main topic in the following pages on the other, I will depart from my terminology in previous studies in which I spoke of (structural) "predicates" and their "complements" (≈ arguments and adjuncts) and instead refer to these units as the TAM/Person-syntagma and the Case-syntagma.[4] The TAM/Person-syntagma corresponds to what I refer to in earlier works (Peterson 2005; 2006; 2007; Peterson and Maas 2009) as structural "predicates" while the Case-syntagma refers to what in those studies is referred to as the "complement of a predicate". The logic of my argumentation has not changed, and this is only a terminological change, as the dual nature of the previous labels may have led some to believe that functional and syntactic categories might not be properly separated. The exact nature of the structure of these two syntagmas from a theoretical perspective will be the topic of section 3 of this study.

The previous terminology was motivated by the fact that what is referred to here as the TAM/Person-syntagma is GENERALLY found in predicative function whereas what was previously referred to as the complement of the predicate, i.e., the Case-syntagma, was QUITE OFTEN an argument or adjunct.[5] However, as I also noted in previous studies, the (structural) "predicate" or TAM/Person-syntagma is not always found in predicative function, while the "complement" or Case-syntagma is by no means restricted to reference.

[4] These terms are motivated by the terms "TAM-syntagm" and the "article-syntagm" for Tongan in Broschart (1997), made to fit the Kharia data.

[5] As this unit is very often not referential, I decided in earlier studies to refer to this syntagma in more general terms, i.e., not as a "referent" (in contrast to the "predicate") but rather simply as the "complement of a predicate".

For example, both the TAM/Person- and the Case-syntagmas may function as modifiers, as in the following examples:[6]

Case-syntagmas

(25) *kuda* *koloŋ* *daru*
millet bread tree
'a millet bread tree' (from a children's story)

(26) *u* *goʈa* *duniya=te* *lebu=ki=ya?* *kahani*
this entire world=OBL person=PL=GEN story
'the story of the people on this entire world'
(Of interest here is that the oblique-case marked *u goʈa duniya=te* modifies *lebu=ki=ya?*.)

TAM/Person-syntagma

(27) *yo=yoʔj* *lebu* *col=ki.*
see=A.PST.1SG man go=M.PST
'The man I saw left' (= literally: 'I=saw man left'.)

Furthermore, the TAM/Person-syntagma may also be used referentially (28), and the Case-syntagma may be used predicatively (29).

(28) *kunɖab aw=ki* *tomliŋ* *khaʈiya* *gam* *ɖom=na* *laʔ=ki=may*
behind QUAL=M.PST milk Kharia say PASS=INF IPFV=M.PST=3PL
ina no *u=ki* *tomliŋ* *uʔɖ=ga* *ɖel=ki=may.* [MT, 1:180]
because this=PL milk drink=FOC come=M.PST=3PL
'[Those who] were in the rear (= literally: 'they were behind') were called "Milk Kharia" because they came drinking milk.'

(29) "… *ro* *u=ga* *ho* *jinis=aʔ* *komaŋ*." [AK 1:57]
and this=FOC that animal=GEN meat
'" …and this [is] that animal's meat."'

2.3 The structure of the Case-syntagma

The structure of the Case-syntagma in Kharia is given in Diagram 1, where non-obligatory components appear in parentheses. The Kleene star "*" denotes that any number of lexemes – including zero – is possible. Note: "lex=gen" refers to a genitive attribute, whose structure can actually be quite complex. This topic will be discussed in detail in section 3.3.

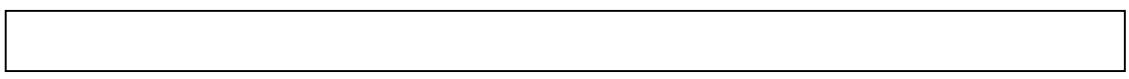


Diagram 1: The structure of the Case-syntagma

[6] Note that there is no evidence for considering the form in (25) to be a compound; cf. Peterson (2011: 118-121).

This structure can be simplified considerably as in (30), where "X" can be any type of semantic base, i.e., it can be a simple lexeme, as in example (1a), or complex, as in (20)b. An overview of the "case" markers is provided in (31), although as we shall presently see, the genitive is best not considered a case in the sense that this term is used in the present study.

(30) X=CASE

(31) Case: Direct (zero marking) – the case of subjects and indefinite direct objects;
Oblique (marked by *=te*) – marks definite direct objects, "indirect objects" and adverbials
Genitive (marked by *=(y)a?*) – the "attributive" case, through which one semantic base is incorporated into a larger semantic base.

Kharia also has three numbers:

(32) Unmarked – both semantically and morphologically, although this usually yields a singular interpretation
Dual – marked by the enclitic *=kiyar*
Plural – marked by the enclitic *=ki*

Note that the genitive "case" has a very different status than the other two given in (31): Whereas the direct and oblique cases are relevant at the clausal level, e.g., marking the relation of a constituent to the predicate (roughly: subject/non-subject), the genitive only plays a role within the semantic base itself (i.e. the "X" in (30)); it allows what without the genitive marker could function as the semantic base of a Case-syntagma (but also of a TAM/Person-syntagma, see e.g. (34)) to function as an attribute within a larger semantic base, a use which is similar to the "adnominal case" in languages with nouns and verbs.[7] Furthermore, the oblique case cannot appear in a TAM/Person-syntagma, as (33) shows. This is not true of the genitive, however, as examples such as (16) above show, repeated here as (34):

(33) **sahar=te=ki=ɲ.*
city=OBL=M.PST=1SG
'I went to the city.'

(34) *ayo=ya?=yo?*
mother=GEN=A.PST.1SG
'he or she made [it] mother's' (lit.: "he or she mother's-ed [it]")

Postpositions[8] behave in this respect similarly to the oblique case: The presence of a postposition is not compatible with any of the elements which mark a TAM/Person-

[7] However, with very few exceptions, the use of the genitive in these environments always appears to be optional, cf. e.g. once again the structure in (25), where the genitive is not used to integrate the "nouny" modifiers into the semantic base. For further details, cf. Peterson (2011: 146-148).
[8] Or more exactly, postpositions and the one preposition in the language, *enem* 'without'. Here and in the following, I will refer to these collectively as "postpositions" for ease of reference.

syntagma, i.e., TAM and subject marking (see next section). In other words, a clause-level unit is either a Case-syntagma or it is a TAM/Person-syntagma, but it cannot be both simultaneously.

Furthermore, a postposition may never appear in conjunction with the oblique case marker, suggesting that they fulfill similar functions. On the other hand, postpositions often take a genitive-marked Case-syntagma: The general rule is that postpositions in Kharia take the genitive case if the reference of the semantic base is definite, or no case marking in the case of indefinite reference.[9] Consider the following examples:

(35)	a.	*ho=kaɽ=a?*	*tay*	b. *kinir*	*tay*
		that-SG.HUM=GEN	ABL	forest	ABL
		'from him/her'		'from (a / the) forest'	

Thus, although it is often convenient to consider the genitive a case, it has a very different status than either the oblique case or the postpositions. On the other hand, the postpositions behave similarly to the oblique case marker and mark the relation of a constituent to the predicate (whether the predicate is formed by a TAM/Person- or Case-syntagma), i.e., as an adjunct. Hence, in the following discussion, the term "case" will refer to the oblique marker *=te*, the postpositions, and – by default – the direct case, with its zero-marking.[10] The genitive, on the other hand, is not a case at all in this sense but is found within the semantic base of either a TAM/Person- or a Case-syntagma.

As noted above, the Kleene star "*" on LEX in Diagram 1 denotes that any number of contentive morphemes, including zero, can appear within the semantic base of the Case-syntagma. As contentive morphemes which denote or can denote properties do not differ in any respect from other contentive morphemes with respect to flexibility (cf. example (6) above), there are no reasons from a structural perspective to assume the presence of adjectives in Kharia. Instead, we merely find any number of contentive morphemes appearing within the semantic base, where each contentive morpheme modifies those to its right. Whether or not these translate as nouns, as in example (25), is of no importance, as there is no reason to differentiate these lexemes on structural grounds from lexemes which translate as adjectives.

There is a slight twist here, however, as the form used in the Case-syntagma is a derived stem, the masdar, which is formed as follows: Polysyllabic stems (= root and derivational markers such as the causative) undergo no changes, while monosyllabic stems are reduplicated. Thus this type of reduplication is purely phonologically motivated, and all contentive morphemes have a corresponding masdar form. Some examples:

(36)	Stem = simple root	Masdar
	kersoŋ 'marry'	*kersoŋ*
	likha 'write'	*likha*
	osel 'become white'	*osel*
	aw 'live, stay, remain'	*aw-aw*
	ru? 'open'	*ru?-ru?*

[9] The role of definacy is actually somewhat more complex, as some postpositions tend to take a genitive-marked complement more often than others. This does not affect our discussion, however, as it is still definacy which accounts for the variation of marking with the individual postpositions.

[10] "By default" here simply means that the respective clause-level unit is followed neither by an overt case marker nor by TAM/basic voice and subject marking. I.e., when no other grammatical marking follows, a clause-level unit can only be interpreted as a Case-Syntagma in the direct case.

soŋ 'buy'		*soŋ-soŋ*
yo 'see'		*yo-yo*

(37)

Complex stem	Masdar
ob-ru? 'have s.o. open' [CAUS-open]	*ob-ru?*
ob-soŋ 'sell' [CAUS-buy]	*ob-soŋ*

(38)

konon bhai	'little brother'	Cf. *konon* '(become) small'	Masdar: *konon*	
maha daru	'the big tree'	*maha* '(become) big'	*maha*	
ru?-ru? ka?bʈo	'open door'	*ru?* '(become) open'	*ru?-ru?*	
rusuŋ o?	'red house'	*rusuŋ* '(become) red'	*rusuŋ*	

The following example, in which two masdars (underlined) appear together, clearly shows the purely phonological nature of this formation, as only the monosyllabic stem reduplicates.

(39) *biru raij aniŋ=a? purkha=ki=ya? ɖoko ro dho?-dho?*
Biru kingdom 1PL.INCL=GEN ancestor=PL=GEN settle and take-RDP

ʈhãʈo hoɖom=a? ti?j=te co=na lam=tej. [Kerkeʈʈā 1990:6]
place other=GEN side=OBL go=INF want=A.PROG
'Biru Kingdom, the place where our ancestors settled and [which they] took, is about to be conquered (= is wanting to go to the side of another).'

As argued in Peterson (2007) and Peterson and Maas (2009), the masdar results from an earlier bisyllabic constraint in Kharia, which itself was the reflex of an older, presumably pan-Austro-Asiatic bimoraic constraint (cf. Anderson and Zide 2002). This bisyllabic constraint in Kharia essentially required any phonological word to be at least bisyllabic. As a unit functioning as the Case-syntagma can be entirely unmarked for case (i.e., the direct case), inalienable possession and number, this would result in a monosyllabic word if the stem were monosyllabic, as there would be no enclitic grammatical marking with which to form a bisyllabic phonological word. It was precisely in this environment that the monosyllabic stem reduplicated (and still reduplicates).[11] As we shall see in the following section, however, in the case of the TAM/Person-syntagma, this was never an issue, as at least one marker for tam and usually also an overt marker for person/number was present, thereby ensuring bisyllabicity even in the case of a monosyllabic stem.

However, the masdar has developed further and is now an independent stem type which, like all stems, can be used not only attributively and referentially but also predicatively in the "generic middle" construction to denote a habitual or iterative action.

---

[11] This is not to say that there are no monosyllabic phonological words in modern Kharia: To begin with, monosyllabic stems borrowed from Indo-Aryan languages never reduplicate, with one exception: *col* 'go', which behaves phonologically the same as the indigenous root *ɖel* 'come' in this and other respects as well (cf. Peterson and Maas 2009: fn. 28). Furthermore, there is an exceptionally small number of indigenous stems, comprising less than 5% of the total indigenous lexicon, which also do not reduplicate in these environments, all highly common forms, such as *laŋ* 'tongue', *ti?* 'hand', *o?* 'house', etc. These would appear to be especially archaic forms, which can be explained by their presumably relatively high frequency. Cf. Peterson and Maas (2009: fn. 31) for further discussion.

Consider first (40), which shows the lexeme *biʔḏ* 'pour out' used in the unmarked construction, where it is restricted to the active voice and is underspecified with respect to habituality.[12]

(40) *iɲ* *ḏaʔ* *biṯh=oʔj*. Unmarked construction
1SG water pour.out=A.PST.1SG
'I poured water out.'

(41) on the other hand has the masdar as its stem and is restricted to the middle voice. This results in a habitual interpretation.

(41) *iɲ* *ḏaʔ* *biʔḏ-biʔḏ=ki=ɲ*. Generic construction
1SG water pour.out-RDP=M.PST=1SG
'I used to pour water out (e.g., that was my job, so I did it all the time).'

The masdar is thus not a noun or adjective, but merely a stem which is positively marked as being what Borer (2005) refers to as [-quantified], i.e., events/states lacking any type of boundary (e.g., generic, atelic, etc.). Although this typically includes nouns and adjectives in languages that have these classes, this is not necessarily so, and Kharia provides positive evidence that this category is not inherently referential or attributive.

In summary, although the masdar at first glance appears to be a kind of derived noun or adjective, it is in fact merely a (derived) stem and, like all stems in Kharia, can also be used predicatively without any further derivational morphology or light verb. It therefore is not proof for the presence of nouns or adjectives in Kharia.

2.4 The structure of the (finite) TAM/Person-syntagma[13]

The overall structure of the fully finite TAM/Person-syntagma in Kharia is given in (42), where "X" has the exact same potential structure as the semantic base of a Case-syntagma, discussed in the previous section, i.e., it can be simple, as in (43), or complex, as in (14)-(19) above. On the status of the facultative "v2s" see further below, beginning with the discussion preceding (46).

(42) X (V2) **(=PERF)=TAM/VOICE=PERS/NUM/HON**

(43) *col=ki=may*
go=M.PST=3PL
'they went'

Note that "X" may also contain marking for inalienable possession (44) and number (45), so that these are not inherently "nominal" categories in Kharia: [14]

[12] Before the past marker *=oʔ*, a stem-final consonant is devoiced and aspirated, hence the stem *biʔḏ* is realized regularly as *biṯh* before *=oʔ*.

[13] For a perspective on the relation of the TAM/Person-syntagma in Kharia to Munda predicative systems in general, see Anderson (2007).

(44) *boksel=nom* *go^{ʔ}ɖ=ki*
sister.in.law=2POSS C:TEL=M.PST
'she became your sister-in-law' (= 'she "your sister-in-law-ed"')

(45) *ho=jeʔ* *u=jeʔ=ki* *go^{ʔ}ɖ=ki*
that=SG.NHUM this=SG.NHUM=PL C:TEL=M.PST
'that became these' (= 'that "these-d"')

Table 1 provides an overview of the TAM/Person-syntagma-defining markers denoting the basic tam-categories, i.e., any finite TAM/Person-syntagma in Kharia is obligatorily marked for one of these categories. As mentioned above, virtually all of these forms are enclitic, denoted here by the sign "=". Only the optative is not enclitic but rather is both a phonologically and syntactically independent word. Furthermore, with the exception of the unmarked 3rd person, which is semantically and morphologically unmarked (although generally interpreted as singular),[15] a finite TAM/Person-syntagma is always overtly marked for person/number/honorific status. Table 2 presents an overview of these forms.

	Active	**Middle**
(Simple) Past (PST)	*=oʔ*	*=ki*
Present (PRS)	*=te*	*=ta*
Present Progressive (PROG)	*=te^{ʔ}jɖ*	*=ta^{ʔ}jɖ*
Irrealis (IRR)	*=e*	*=na*
"Past II" (PST.II)[16]	*=khoʔ*	
Perfect (PERF)	*=siʔ(ɖ)*	
Optative (OPT)	*guɖuʔ / guɽuʔ*	

Table 1: The basic TAM/voice categories

	Singular	**Dual /HON**		**Plural**	
		Inclusive	**Exclusive**	**Inclusive**	**Exclusive**
1	*=(i)ɲ*	*=naŋ*	*=jar*	*=niŋ*	*=le*
2	*=(e)m*	*=bar*		*=pe*	
3	–	*=kiyar*		*=ki / =may*	

Table 2: Markers for Person/Number/Honorific status

The "v2s" mentioned in (42) are markers of *Aktionsart* (e.g., telicity, conativity, etc.), the passive, benefactive, etc., which derive from one-time contentive morphemes. These units

---

[14] *go^{ʔ}ɖ* 'C:TEL' in these examples is not a verb or verbalizing element but rather a "V2" (see following discussion in main text) denoting telicity. It is also compatible with event-denoting lexemes, such as *ɖoko* 'sit down' in example (47) or *hoy* 'become' in (54).

[15] For the sake of presentation, "zero-marking" is presented in Table 2 as 3rd person, singular. It is important to remember, however, that it is both morphologically AND semantically unmarked.

[16] The "Past II" derives from the past perfect form *=sikh=oʔ*. Its meaning does not seem to differ at all from that of the (simple) past form in *=oʔ*. The only difference to the simple past that I am aware of is that the use of the "Past II" is considered incorrect and frowned upon, especially by older, educated individuals.

follow the semantic base of the TAM/Person-syntagma and precede the enclitic TAM-markers.

(46) *leʔj bay=oʔ*
curse EXCESS=A.PST
'he or she gave [somoneone] a good scolding'

Unlike the enclitic tam-markers, these V2s can be separated from the semantic base of the TAM/Person-syntagma and from each other by the floating pragmatic enclitic markers (*=ga* 'FOC', *=ko* 'CNTR', *=jo* 'add'), which are otherwise always the final element in a linguistic unit.[17]

(47) *ho=kaɽ ɖoko=ga goʔɖ=ki.*
3=SG.HUM sit.down=FOC C:TEL=M.PST
'He or she sat down.'

Finally, two or more TAM/Person-syntagmas which would have identical TAM and PERS/NUM values in their fully finite forms may share one or both of these, i.e., either PERS/NUM or both TAM and PERS/NUM may be elided in the non-final predicates. This also demonstrates yet again the enclitic nature of these markers. Compare the following examples, where shared markers are underlined:

(48) *tar ol=e=pe*
kill bring=A.IRR=2PL
'kill [the animal and] bring [it back]!'

(49) *tar oʔ-gur=e=niŋ*
kill CAUS-fall=A.IRR=1PL.INCL
'we will kill the animal' (lit.: "will kill [it and] cause [it] to fall")

(50) *ɲog=e uɖ=e=kiyar*
eat=A.IRR drink=A.IRR=DU
'They two will eat and drink'

2.5 TAM/Person- and Case-syntagmas in Kharia: A summary

In sum, although we could not go into this issue here in detail, various criteria were mentioned in the preceding pages which, taken together, strongly suggest that Kharia does not possess nouns and verbs as lexical categories, among them:

- All contentive morphemes in Kharia are extremely flexible (cf. e.g. examples (1)-(12)).
- Assuming the presence of nouns and verbs in Kharia brings with it a number of additional problems: As examples such as (14)-(19) show, assuming the presence of

[17] On the somewhat ambivalent status of the V2s as phonological words, see the discussion in Peterson and Maas (2009, §6).

lexical classes in Kharia would force us to productively zero-derive verbs not from nouns but rather from entire NPs, i.e., not in the lexicon but in the syntax.

I would like to stress at this point once again that I do not wish to claim here that it is impossible to analyze (Simdega) Kharia as a language with familiar parts of speech such as noun, verb and adjective. This is perhaps possible, at least if "rampant zero derivation" (Evans and Osada 2005) is assumed and if one is not bothered by zero-derivation applying regularly to clause-level units or "phrases". Instead, I merely argue that it is most economical to assume that Kharia simply possesses TAM/Person- and Case-syntagmas, both of which consist of two obligatory parts: The first part contains semantic information which serves to identify the entity, action, or state. Crucially, this element has the same potential structure for both TAM/Person- and Case-syntagmas. The second part of the syntagma contains purely grammatical information such as case, *Aktionsart*, TAM and person/number/honorific status and it is this element alone which distinguishes TAM/Person-syntagmas from Case-syntagmas.

As to my knowledge it is one of the most basic tenets of (descriptive) linguistics that only those categories should be assumed for a particular language or construction for which there is clear POSITIVE evidence, I do not assume lexical parts of speech such as noun and verb, which perhaps can be made to fit the data but which do not seem to be necessary, except under certain theoretical assumptions.

Having said that, I nevertheless believe that there is also very suggestive evidence that an analysis of Kharia as containing nouns and verbs, even if possible, will eventually run into serious problems if it is applied consistently. As noted in section 2.3, it is not possible in Kharia to combine TAM/person marking with case marking (other than the genitive) or with postpositions. Thus, a construction such as that in (51), which ends in a postposition, cannot be used predicatively by adding TAM/person marking, as (52) shows.

(51) *iku^{ʔ}ɖ jughay du?kho buŋ*
very much sorrow INST
'very sad', literally: 'with very much sorrow'

(52) **iku^{ʔ}ɖ jughay du?kho buŋ go^{ʔ}ɖ=ki.*
very much sorrow INST C:TEL=M.PST
'(He) became very sad' (= 'with very much sorrow')

However, if the semantic part of this unit is re-ordered, so to speak, such that we now have two different "phrases", as in (53), instead of just the one in (51), the unit which is not marked by a postposition may serve as the semantic base of a TAM/Person-syntagma, giving the impression that the parts of the original constituent have traded places. This is a common predicate type in Kharia, especially with "psyche predicates" (cf. Peterson, 2011: 111; 220).

(53) *gupa lebu du?kho buŋ iku^{ʔ}ɖ jughay go^{ʔ}ɖ=ki.*
watch person sorrow INST very much c:TEL=M.PST
'The shepherd (= watch person) became very sad.' [RD, 1:18]

Although this construction raises a number of interesting theoretical questions, what is relevant to our discussion here is the fact that the unit ending in the postposition may not be directly followed by a TAM marker, whereas the unit which does not end in a

postposition/case marker can, as is to be expected. Crucially, however, if a change-of-state marker such as the "light verb" *hoy* 'become', borrowed from Kharia's Indo-Aryan neighbour Sadri, is inserted between *buŋ* 'INST' and *go^{ʔ}ɖ=ki* in (52), this results in a perfectly grammatical predicate, consisting of a Case-syntagma and a marker of qualitative predication (≈ copula).

(54) *iku^{ʔ}ɖ jughay duʔkho buŋ hoy go^{ʔ}ɖ=ki.*
very much sorrow INST become C:TEL=M.PST
'(He) became very sad.'

This strongly suggests that assuming hidden verbs is problematic in Kharia, as these would have to behave differently than the "light verb" *hoy* 'become' or for that matter any other supposed verb, since they would only be found following "predicate nominals" which are not case marked, crucially precisely those units which are considered here to be the semantic base of either a TAM/Person- or a Case-syntagma. Thus, assuming hidden verbs does not simplify the analysis at all but rather complicates it considerably: They are motivated solely by the theorectical assumption that every language must have nouns and verbs but have no explanatory value and also behave differently from all other supposed "verbs".

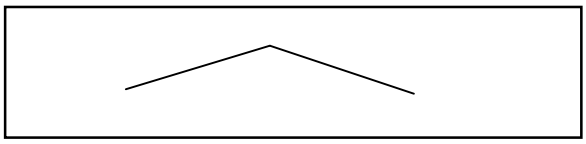
In sum, in my view (Simdega) Kharia presents a number of arguments against assuming the presence of nouns, adjectives and verbs – hidden or otherwise – but no evidence in their favor, hence an analysis which does not have recourse to these categories is to be preferred over an analysis which assumes their (overt or covert) presence. Whether or not it is POSSIBLE to assume their presence is not at issue here. Rather, the remainder of this study assumes that these categories are not essential for a description of Kharia and attempts to formalize this assumption in such a way that this analysis can easily be implemented in any (non-dogmatic) monostratal approach.

3 A formal account

In the following pages I describe the structure of TAM/Person- and Case-syntagmas in Kharia in detail. The discussion presented here is deliberately theory neutral and makes no assumptions other than "monostratality", i.e., that observed language is not derived from some underlying level, although even this may not turn out to be a necessary requirement. As such, it can easily be applied – with relatively minor changes – to any non-derivational theory, such as e.g. Role and Reference Grammar (RRG), Lexical-Functional Grammar (LFG), Head-driven Phrase Structure Grammar (HPSG) and others.

3.1 Co-heads

As argued in the preceding sections, assuming the presence of nouns and verbs in Kharia leads to a number of problems for which there appears to be no easy solution. For this reason I have opted to do away with these concepts altogether and assume only TAM/Person- and Case-syntagmas. In this section, I present a formalized alternative to the traditional parts of speech in which TAM/Person- and Case-syntagmas in Kharia both consist of two parts, a semantic base or "head", which I will refer to as the content head, and a functional head, i.e., we have two "co-heads". While the content head has the same



potential form for both TAM/Person- and Case-syntagmas, it is the functional head which differentiates between the two structures. This is shown schematically in Diagram 2.[18]

Diagram 2: Common structure of TAM/Person- and Case-syntagmas in Kharia

The idea of co-heads derives from LFG (cf. e.g. Simpson 1991: 72-83; Bresnan 2001: 109; Falk 2001: 39; 83) and merely means that any construction, for example an NP in English, can have more than one head. Consider e.g. the following phrase-structure rule for the English NP from Simpson (1991: 79):

(55) NP → (Det) (Adj) N
↑=↓ ↑=↓

(55) states that there are two co-heads in the English NP (or DP, depending on one's perspective), marked here by the symbol "↑=↓". These two heads are usually referred to as the functional head (= Det) and the lexical head (= N). However, as the element in Kharia which most closely corresponds to the English head noun, or lexical head, need not always be overt (cf. example (20)b), I deviate here somewhat from standard terminology and use the term "content head" instead of the more usual "lexical head", although the function of the two is virtually the same since this unit carries semantic (and referential) information, as opposed to purely functional information.

The phrase-structure rule given in (55) fits the structure of English quite well, as both the lexical and functional heads are separate words. In fact, this is a prerequisite for any such analysis in theories such as LFG which make strong assumptions with respect to the status of the "word": Here a word is a unit which emerges from the lexicon intact and which is opaque to syntax. This principle is generally referred to as the "Lexical Integrity Principle" and basically means that if it cannot be shown for a certain language that, e.g. the lexical and functional parts of an NP are "words", then the analysis above for the English NP will not be directly transferable to that language. For languages such as e.g. Hebrew, in which definacy is expressed through a prefix and not a separate word, such an analysis will have to be modified, as only words occupy syntactic nodes (cf. e.g. the discussion in Falk (2001: 40f)).[19]

---

[18] Note that, if one prefers to implement the present analysis into an HPSG format, it will be necessary to assume that one of these elements is the "head" and that this is a headed phrase. In this case, since it is the functional head which determines the status of the unit as a whole and since the semantic head has the same potential form for both syntagmas, it would clearly be preferable to consider the functional head to be the head of the construction.

[19] It is of course debatable whether the Lexical Integrity Principle is really necessary, as several theoretical frameworks, such as e.g. Role and Reference Grammar (RRG), do not make such requirements. However, as LFG to my knowledge makes the most stringent requirements with respect to this principle, and as the principle CAN be applied to Kharia with no further difficulty since virtually ALL grammatical markers are enclitic, I will adhere to this principle in the present study and will not deal further with its (un-)necessity in general. For theories such as RRG which do not require this principle, the following discussion will simply be irrelevant.

The remainder of this section is structured as follows: After addressing the issue of lexical integrity in section 3.2, we deal in 3.3 with the structure of the Case-syntagma. Following this, section 3.4 deals with the structure of the TAM/Person-syntagma.

3.2 "Words" revisited – the question of lexical integrity

As noted above, the principle of lexical integrity is central to LFG and hence must be addressed if the following proposal is to be applicable to LFG as well as other, less restrictive theories. The definition of this principle which will be assumed here is as follows:

(56) **Lexical Integrity Principle** – Morphologically complete words are leaves of the structure tree and each leaf corresponds to one and only one structure node. (Cf. e.g., Bresnan 2001: 92; Falk 2001: 26)

That is, only "words" can occupy a node in the constituent structure, not parts of words, and the analysis of syntax must always be in terms of whole words, not their internal structure.

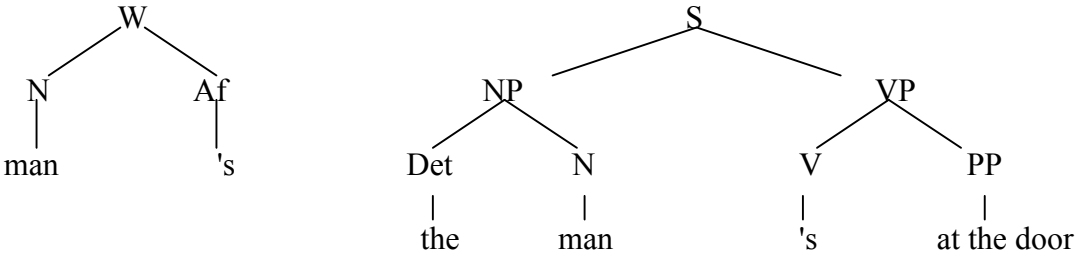
The question immediately arises as to just what is meant by the term "word", as both phonological and syntactic aspects (among others) must be taken into account, i.e., there are phonological and syntactic words. In languages such as English and undoubtedly in most fusional languages such as Latin or Greek, these two types of words will generally coincide and only few morphemes, such as English *'ll* (*will*), *'s* (*is; has*) or the well-known enclitics of the classical languages such as Latin *que* 'and' do not have a perfect overlap. The general stance on this issue within mainstream LFG can be summarized as follows: What matters for a syntactic description of the clause/sentence is not whether we are dealing with phonological words but whether we are dealing with what may be termed syntactic words or "atoms" (following Di Sciullo and Williams 1987; cf. also Falk 2001:6).

It was argued in section 2.1 that, with the exception of the few derivational markers found in Kharia such as the causative marker, all remaining grammatical markers are either enclitics or syntactic and phonological words, such as postpositions or the v2s. Concerning the status of clitics, there is general agreement that clitics are syntactic atoms, although they are phonologically not independent words. This is also true in Kharia, as these elements are both what are referred to as "phonological clitics" as well as "morphosyntactic clitics" by Anderson (2005). We thus assume here that clitics are "words" in the sense of the Lexical Integrity Principle which occupy their own nodes in the structure.

By way of example, consider the simple English sentence *The man's at the door.* Loosely following Sadock's (1991) terminology, there is a "mismatch" here between phonology[20] and syntax. This can be demonstrated by the two diagrams in (57) from Sadock (1991: 50, "W" = word, "S" = sentence, "Af" = affix).

[20] Actually, Sadock speaks here of "morphology", as he is discussing morphological words. Nevertheless, his arguments also hold if this example is viewed in terms of phonology, as his "morphological" word, *man's*, is also a phonological word.

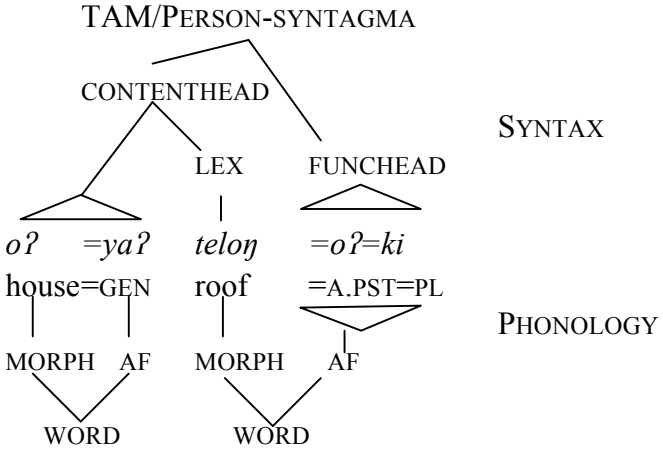
(57) *The man's at the door.*



In other words, *'s* forms part of the phonological word *man's*, although it is simultaneously the V in the VP *[i]s at the door.*

The same principles can be applied to enclitics in Kharia. Consider (58): Although the details of this structure have not yet been dealt with, the labels given in the upper half of the diagram should be intuitively clear enough for the moment. Each of the enclitic markers is considered a syntactic atom while at the same time forming a phonological word with its host.

(58)



'They roofed the house / put a roof on the house.'

The realization that these enclitic markers are syntactic atoms and not suffixes has an additional advantage: It has often been claimed that, since the internal structure of words is opaque to syntax, certain processes cannot refer to their internal structure. Cf. e.g. the following quote from Simpson (1991: 56):

> "This generalisation can be captured on three assumptions, first, that words are opaque not only to movement and deletion processes, but to all other syntactic processes, second, that anaphoric reference and question-formation require coindexation and third, that coindexation is a syntactic process. From these assumptions it follows that parts of words cannot be anaphoric: **Fred is a Volvo-lover, but Bill is a them-hater/it-hater.*"

Now consider once again examples (44) and (45), repeated here as (59) and (60):

(59) *boksel=nom* *goʔɖ=ki* . vs. *boksel-nom-goʔɖ-ki*
sister.in.law=2POSS C:TEL=M.PST
'She became your sister-in-law.'

(60) *ho=je?* *u=je?=ki* *go?ɖ=ki.* vs. *ho-je?* *u-je?-ki-go?ɖ-ki*
that=SG.NHUM this=SG.NHUM=PL C:TEL=PL
'That became these.'

If we view the grammatical markers in these examples as affixes, as has traditionally been done, this results in a number of theoretical problems (at least for some researchers), since e.g., *nom* in (59) would be a referential part of a word, with similar comments holding for *u-je?-ki* in (60). On the other hand, if it is recognized that these grammatical markers are not suffixes but rather enclitics, which we must for independent reasons (cf. section 2.1), and that V2s such as *go?ɖ* are words in both a phonological and syntactic sense, then they are syntactic atoms and no problems arise, as the anaphoric/deictic elements are words from a syntactic perspective.[21]

The remainder of this chapter is dedicated to explicating the exact structure of the content and functional heads discussed above.

3.3 The Case-syntagma

Diagram 1 from section 2.3, repeated here as Diagram 3, shows the structure of the Case-syntagma in Kharia, with non-obligatory components in parentheses.

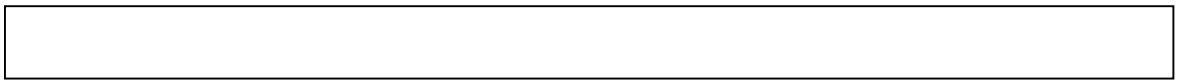


Diagram 3: The structure of the Case-syntagma

The following phrase structure rules follow directly from Diagram 3, in which the semantic base, i.e., the entire structure to the left of CASE, will be referred to as the content head (= conthead). These rules will be discussed and modified somewhat in the following pages.

(61) CASE-SYNTAGMA → CONTHEAD CASE
(62) CONTHEAD → (LEX=GEN) (DEM) (QUANTP) (LEX=GEN) (LEX*) (POSS) (NUM)
(63) QUANTP → QUANT (CLASS)

Recall from section 2.3 that the only element which is obligatory in Diagram 3 is case. Although at least some unit from the structure preceding case marking must be present so that there is some kind of semantic base, any of the elements preceding case can serve as the content head of the Case-syntagma, with the following exceptions: classifier cannot appear alone, as it requires a QUANTifier and so can only appear if a quantifier is present. Similarly, POSS and NUM, themselves enclitic, require a phonological host to which they can attach. Furthermore, CASE is the only element in this structure which cannot appear in a TAM/Person-syntagma, as (33), repeated here as (64), shows.

(64) **sahar=te=ki=ɲ.*
city=OBL=M.PST=1SG
'I went to the city.'

[21] I stress here once again that this discussion is not intended as a discussion on the necessity of assuming the validity of the Lexical Integrity Principle in general but rather, if this principle is assumed to be necessary, the analysis proposed here will nevertheless still be valid.

Clearly, CASE is the functional head of the Case-syntagma, as there would otherwise be no reason why it should not be able to appear in a TAM/Person-syntagma, whereas the remaining elements are not only facultative but may appear in the content head of both Case-syntagmas and TAM/Person-syntagmas. As (59)-(60) show, this also applies to POSS and NUM.

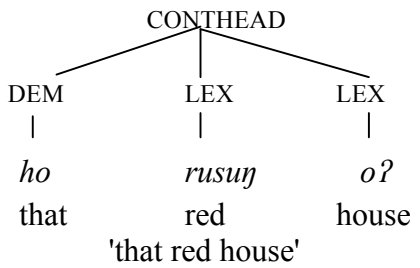
Recall that CASE applies only to the oblique marker *=te* and the postpositions such as *tay* 'from; ABL', *thoŋ* 'for; PURP', etc., as well as by default to the "zero-marker" of the direct case, i.e., those elements which signal the role of the Case-syntagma in a clause. However, it does not apply to the genitive: As (65) shows, the genitive can occur in the content head of both Case-syntagmas and TAM/Person-syntagmas and is therefore not a functional head.

(65) *ayo=ya?=yo?j.*
mother=GEN=A.PST.1SG
'I made [it] the mother's / gave it to the mother.'

The structure suggested here thus allows us to conveniently divide the Case-syntagma in Kharia into two parts, both of which have a distinct function: The content head provides semantic information, while the functional head, as the name implies, designates the function of this unit in a particular clause. The functional head of the Case-syntagma is not always overtly expressed, since zero-case marking, or rather the lack of any overt case marking, signals the direct case. As such, the lack of any overt functional head unambiguously marks a unit as a Case-syntagma in the direct case.

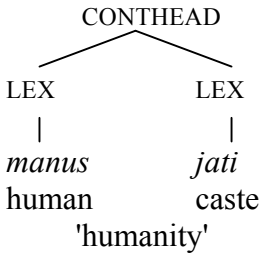
The Kleene star "*" on LEX in Diagram 3 indicates that any number of lexemes is allowed, including zero: There can be none, as in *ho=ki=te* [that=PL=OBL] 'those (OBJECT); them', where the supposed lexical head is ommitted as its identity is either apparent from context or is considered unimportant. There can of course also be one LEX, as in example (1a), but there may also be more than one lexeme found in the semantic base, in which case non-final lexemes modify those lexical units which appear to their right. Recall from section 2.2, e.g. example (25), that I do not distinguish structually between any kind of lexical head and its modifiers: Although there may be reasons for doing so in a semantic analysis, there is no reason to distinguish between the two on structural grounds in Kharia. Thus, in the following two examples, I have neither evidence for the presence of an adjective in (66) nor that (67) is a compound.[22] Rather, both simply contain two LEXs.

(66)



[22] Cf. footnote 6.

(67)

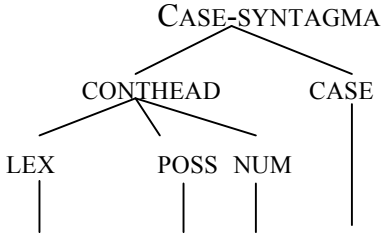


Finally, the content head may consist of a single pro-form, PRO. As this proform is never modified by a demonstrative, genitive phrase or quantifier,[23] I will consider it to be the only element of the content head as this has been defined here. The following summarizes the phrase-structure rules for Case-syntagmas presented so far:

(68) CASE-SYNTAGMA → CONTHEAD CASE[24]
(69) CONTHEAD → ((LEX=GEN)(DEM) (QUANTP) (LEX=GEN) (LEX*) (POSS) (NUM))v PRO
(70) QUANTP → QUANT (CLASS)

In (71)-(73) I present structures for a few Case-syntagmas to illustrate the structure discussed above.

(71)



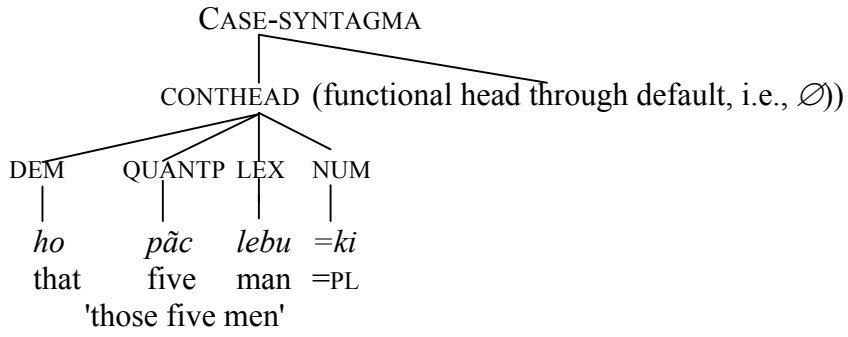
konselɖu? *=ɖom=ki* *=te*
woman =3POSS=PL =OBL
'their wives (= women)' (e.g., as the object of the clause)

[23] It can, however, be modified for some speakers by a clausal modifier ("relative clause"). See Peterson (2011: 406-407) for details. For the sake of brevity, clausal modifiers will not be dealt with here.
[24] For reasons of space, we will not deal in this study with coordinated structures such as that in (23), repeated here as (i), where the semantic base essentially consists of several contheads (underlined):

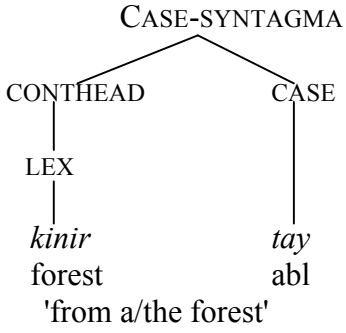
(i) [[[*ayo*] [*aba*]] *ro* [[*boker*] [[*kulam*]]]=*ɖom=ki*
mother father and brother.in.law brother=3POSS=PL
'their mother, father, brother-in-law and brothers'

As this is effect simply requires that CONTHEAD can be recursive and poses no further theoretical problems, I will not pursue this topic further here.

(72)



(73)



Recall that the genitive is not a case marker in the sense that it does not mark the relation of a Case-syntagma to the predicate (whether TAM/Person- or Case-syntagma) of the clause. Rather, it denotes that a possible CONTHEAD is either part of a larger conthead or that a certain postposition takes a CONTHEAD in the genitive. Consider the following example, in which the genitive is used to integrate one potential conthead into a larger conthead:

(74) *oʔ=yaʔ* *teloŋ*
house=GEN roof
'the roof of the house'

The genitive marker in (74) denotes that *oʔ* 'house' restricts the possible interpretations of *teloŋ* 'roof', i.e., what kind of roof is involved, namely that of a house. As the unit to which genitive marking attaches is itself a potential CONTHEAD (here: *teloŋ* 'a/the roof'), we can denote its presence simply by indicating that it is an optional enclitic marker at the end of all other elements which can combine to form the CONTHEAD. Thus, (62) and (69) above can be simplified as follows:

(75) CONTHEAD → (CONTHEAD) (DEM) (QUANTP) (CONTHEAD) (LEX*) (POSS) (NUM) (GEN)
∨ PRO (GEN)

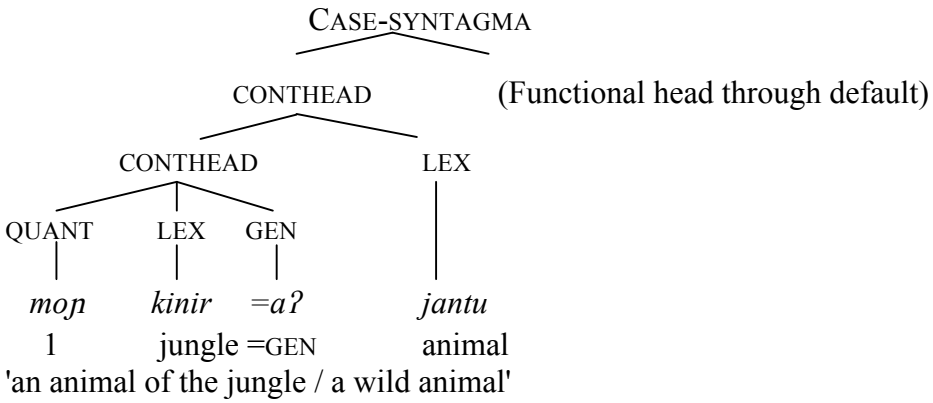
The internal CONTHEADs contained in (75) may – but need not – be marked by the genitive, which now can be seen to have basically the same status as poss and num. As such, the genitive-marked unit should be able to appear wherever a content head unmarked for the genitive may appear, including within the content head of a TAM/Person-syntagma. As example (17) shows, repeated here as (76), this is indeed the case with respect to TAM/Person-syntagmas (which will be discussed in the following section):

(76) a. *iɲ* *ho=kaɽ=te* *iɲ=aʔ=yoˀj.*
1sg that=SG.HUM=OBL 1SG=GEN=A.PT.1SG
'I adopted him/her (i.e., I made him/her mine).'

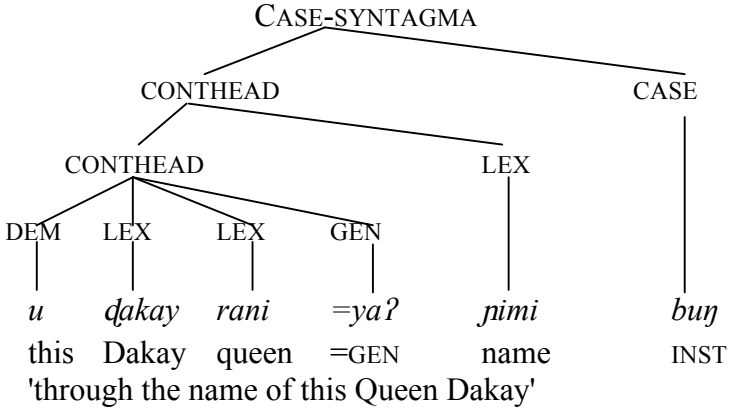
b. *am* *ho=kaɽ=te* *am=aʔ=yoˀb.*
2sg that=SG.HUM=OB 2SG=GEN=A.PST.2SG
'You adopted him/her.'

Making use of this revised structure rule, we can now illustrate the structure of Case-syntagmas with a genitive marker as in (77)-(79).

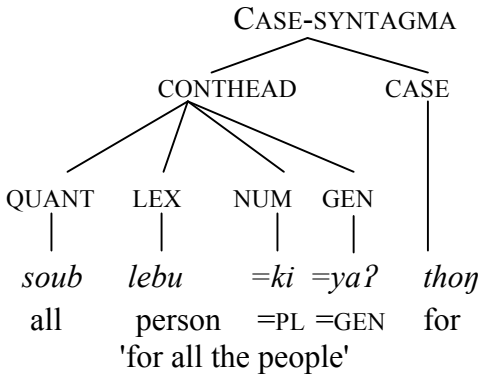
(77)



(78)



(79)



CONTHEAD (V2*) (=PERF)=TAM/VOICE=PERS/NUM/HON

3.4 The TAM/Person-syntagma

The TAM/Person-syntagma similarly consists of a semantic head and a functional head. For the sake of brevity we will restrict our discussion here to the structure found in event-narrating predicates.[2]

Event-narrating predicates contain markers for tam/basic voice and person/number/honorific status following the CONTHEAD, possibly in addition to a "V2" or the perfect marker, as in the following example, repeated from above.

(80) *iɲ ho=kaɽ=te iɲ=a?=yo?j.*
1SG that=SG.HUM=OBL 1SG=GEN=A.PST.1SG
'I adopted him/her (i.e., I made him/her mine).'

The structure of the TAM/Person-syntagma is given in Diagram 4, based on the discussion in sections 2.3 and 3.3.

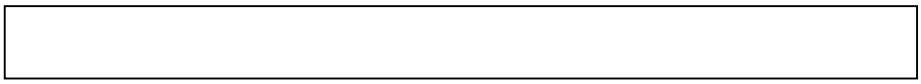


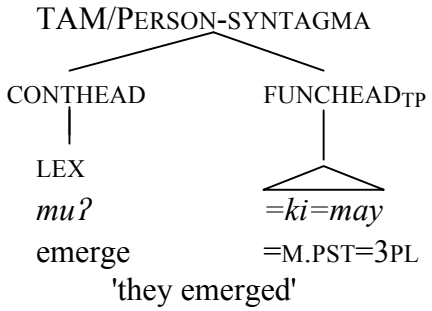
Diagram 4: Schematic overview of the TAM/Person-syntagma

In a first approximation, the TAM/Person-syntagma consists of the same potential semantic base as the Case-syntagma but has a different type of functional head, which we will refer to for the moment simply as $\text{FUNCHEAD}_{\text{TP}}$. This allows us our first formal approximation of the structure of the TAM/Person-syntagma, given in (81)-(82).

(81) TAM/PERSON-SYNTAGMA → CONTHEAD $\text{FUNCHEAD}_{\text{TP}}$
(82) $\text{FUNCHEAD}_{\text{TP}}$ → (V2*) (PERF) TAM/VOICE PERS/NUM/HON

Examples (83)-(87) provide a few examples illustrating this structure. For ease of presentation, the functional markers from rule (82) are presented here as an unanalyzed whole. Their internal structure will be dealt with in section 3.4.1.

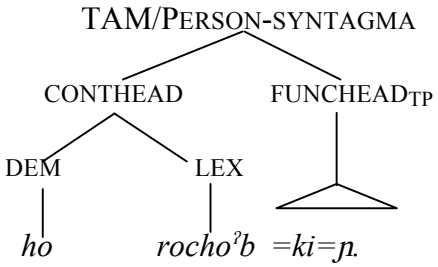
(83)



'they emerged'

[2] Qualitative predicates, which denote the identity of a particular referent or a permanent state (*I am a man, that is true, she is honest*) or refer to a temporary state or position (*the book is on the table*) will not be dealt with here. These predicates consist of a Case-Syntagma and also often a unit translating as a copula in English. As such, their incorporation into the present format is trivial and need not be dealt with here.

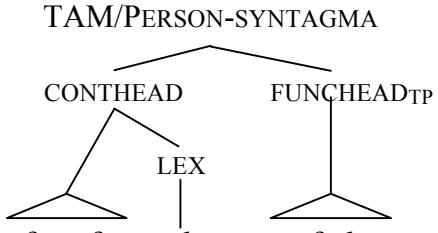
(84)



that side =M.PST=1SG

'I moved to that side.'

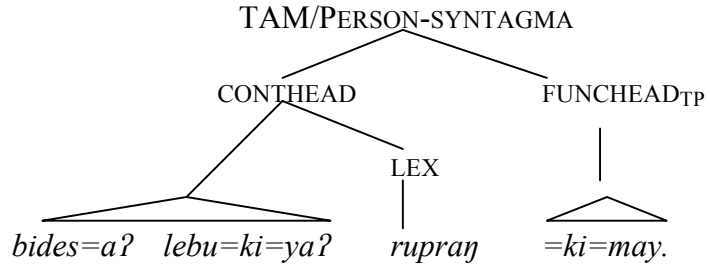
(85)



house=GEN roof =A.PST=PL

'They roofed the house.'

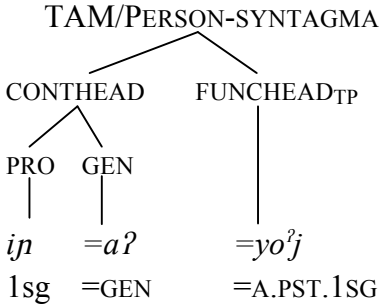
(86)



abroad=GEN person=PL=GEN appearance =M.PST=3PL

'They took on the appearance of foreigners (e.g. by living abroad so long).'

(87)



1sg =GEN =A.PST.1SG

'I adopted [someone]; I made [someone/something] mine'

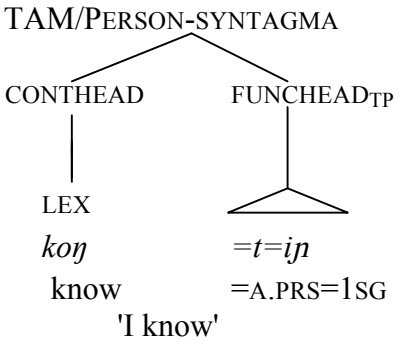
3.4.1 The internal structure of the functional head of the TAM/Person-syntagma

The functional head of TAM/Person-syntagmas is actually more complex than the previous examples suggest and cannot simply be considered an element consisting of TAM-marking + PERS/NUM/HON-marking. This becomes especially apparent e.g. in the case of negation, to which we now turn.

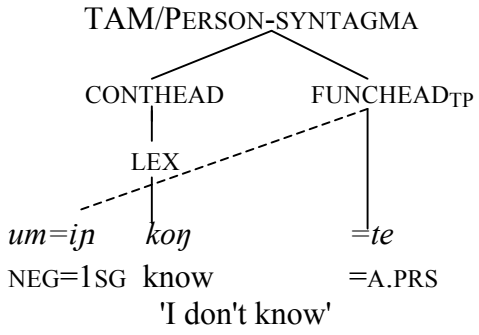
Sentential (indicative) negation in Kharia is expressed through the negative particle *um*. When the TAM/Person-syntagma is negated, PERS/NUM/HON-marking attaches to the negative morpheme *um*, which precedes the now semi-finite TAM/Person-syntagma.[3] Compare the structure diagram in (88) of a non-negated TAM/Person-syntagma with that of the corresponding negated TAM/Person-syntagma in(89). The negative marker *um* is not considered here part of the functional head of the TAM/Person-syntagma, as it can appear both in Case- and in TAM/Person-syntagmas. I will not deal with its status further but will simply treat it here as a clause-level polarity marker with no further comment.

Note that the structure in (89) for the negated TAM/Person-syntagma is not allowed in theories such as LFG (although this is unproblematic in others), as the information which in the present analysis should belong to the node of the functional head of the TAM/Person-syntagma is discontiguous.

(88) ***Non-negated TAM/Person-syntagma***



(89) ***Negated TAM/Person-syntagma***

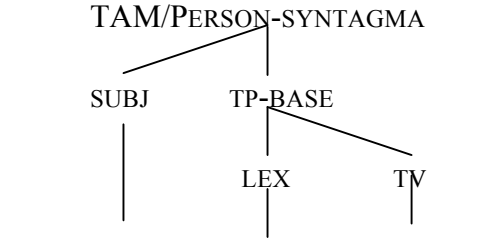


[3] The negative morpheme *um* is not an auxiliary or any type of verb, as one anonymous referee suggested. To begin with, it cannot mark for any tam categories. Furthermore, unlike other persons, subject marking for the 2nd person, singular may attach to either *um* or it may appear at the end of the TAM/Person-syntagma. Hence, if *um* were considered a verb, it would be the only verb to show these two traits. Finally, *um* is compatible with other lexemes denoting states (translating as English adjectives), abstract notions (translating as English nouns), etc., including when these are not TAM/Person-syntagmas. Consider the following few examples: *um lere?* 'unhappy' (*lere?* 'joy; joyous; rejoice'), *um bes* 'bad' (*bes* 'good'), *um dharmi* (*dharmi* 'righteous(ness)') 'unrighteous(ness)', *um suru* 'non-beginning', etc. Hence, neg=pers/num/hon cannot be considered a verb or negative auxiliary but is simply the negative polarity particle followed by enclitic subject marking.

(88) and (89) show that the mobility of the enclitic subject marker is different from that of the TAM/BASIC VOICE markers and that the two are independent of one another. Hence, our phrase-structure rules must be modified to account for this.

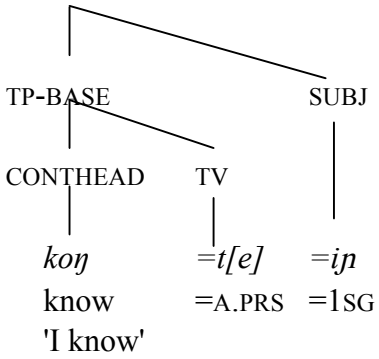
I suggest the structure in (90) for TAM/Person-syntagmas in general, where "TV" stands for "Tense-aspect-mood/basic voice" (= active/middle): In this analysis, a TAM/PERSON-SYNTAGMA obligatorily consists of at least one TP-BASE and SUBJect marking (with "zero-marking" by default for the 3rd person). The TP-BASE then contains the semantic information but also TV.

(90)



The same structure also holds for non-negated TAM/Person-syntagmas, as (91) shows.

(91)



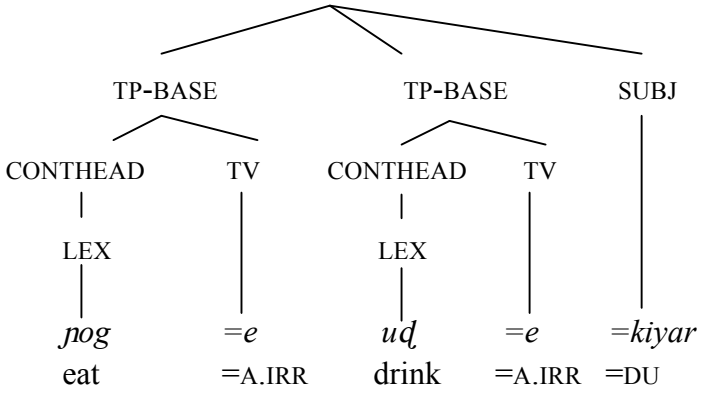
Although this structure may seem rather *ad-hoc*, having been created only to account for the structure of negated TAM/Person-syntagmas, it is also required elsewhere: recall, e.g., from examples (48)-(50) that multiple TAM/Person-syntagmas occurring together – using the terminology developed in this section – MAY share either SUBJ-marking or TV and SUBJ-marking when these (would) have identical values. Consider example (92) (= (50)). Hence, we will assume this structure in general.[4]

---

[4] Such a structure is further motivated by morphologically partially finite forms found in "relative clauses" such as the following, where only SUBJ-marking is elided. Cf. Peterson (2011: 313-317; 410-414) for further discussion.

(ii) *iŋ=te yo=yo?* ***lebu=ki*** *ula? likha=yo?=ki.* Cf. *yo=yo?=ki* [see=a.pt=p] 'they saw'
1sg=obl see=a.pst person=pl letter write=a.pst=pl
'**The people** *who saw me* wrote a letter.'

(92) TAM/Person-syntagma



'They two will eat and drink'

We may summarize the above discussion in the following rules, where the Kleene plus sign "$^{+}$" means that any number of TP-BASEs is allowed, as long as at least one is present (the status of the V2s will be discussed below in section 3.4.2).

(93) TAM/PERSON-SYNTAGMA → TP-BASE^{+} SUBJ
(94) TP-BASE → CONTHEAD TV

TV consists of a tense-aspect-mood (TAM) attribute and a basic-voice (BV) attribute, both defined disjunctively. Each of these two attributes receives a unique value and each combination of TAM and BV also yields a unique value corresponding to a unique morpheme, shown in Table 3.

(95) TAM = PAST ∨ PRESENT ∨ PROGRESSIVE ∨ IRREALIS
(96) BV = ACTIVE ∨ MIDDLE

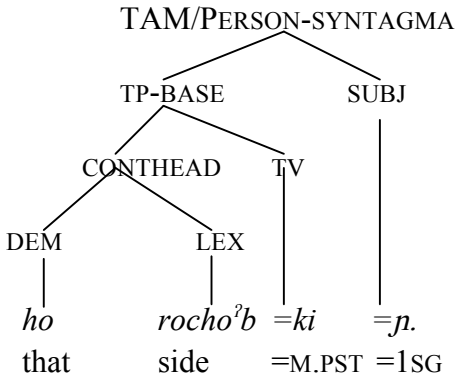
PAST ∧ ACTIVE	*=(y)oʔ*
PAST ∧ MIDDLE	*=ki*
PRESENT ∧ ACTIVE	*=te*
PRESENT ∧ MIDDLE	*=ta*
PROGRESSIVE ∧ ACTIVE	*=teʔj*
PROGRESSIVE ∧ MIDDLE	*=taʔj*
IRREALIS ∧ ACTIVE	*=e*
IRREALIS ∧ MIDDLE	*=na*

Table 3: TAM ∧ BV

We conclude this section with two examples which illustrate this structure with TAM/Person-syntagmas with a complex, "nouny" CONTHEAD.[5]

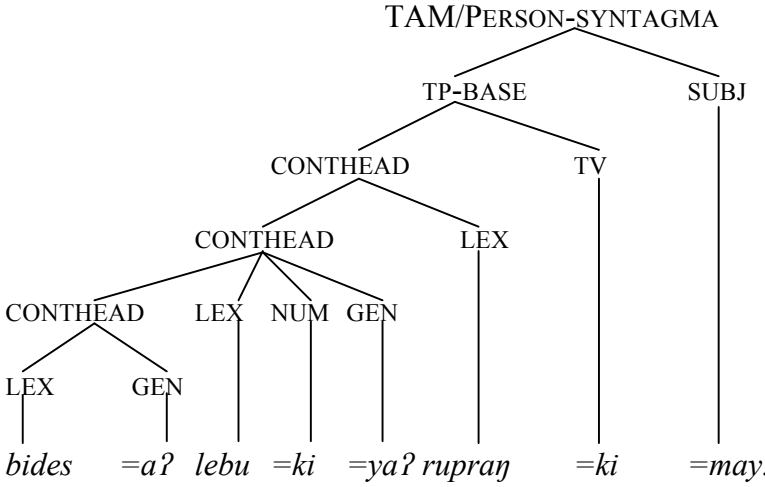
[5] Cf. also the preliminary structures in (84) and (86), which are given in their revised forms here.

(97)



'I moved to that side.'

(98)



'They took on the appearance of foreigners (e.g. by living abroad so long).'

3.4.2 V2s, telicity and the perfect

There is a slight complication with respect to the V2s which we will now deal with. Recall the discussion of the masdar in section 2.3: The masdar does not appear as a separate entity in this formal analysis as it is simply a (derived) lexical stem which, like all roots and stems in Kharia, may appear in referential, attributive and predicative functions.

Nevertheless, the masdar is of importance as it demonstrates that the V2s, which have until now been treated as a single category, in fact do not form a homogeneous class. Consider the following example, which at first glance suggests that the passive/reflexive marker *ɖom* follows the masdar:[6]

(99) *loʔɖho khaɽiya, koɽa, niga oɖoʔ oɖoʔ uʔchi ɖom* ***jati*** ...
later Kharia Mundari Kurukh other REP despise PASS ethnic.group
'Later, the Kharia, Mundari, Kurukh and other *despised* **ethnic groups** …'
[Pinnow 1965:120]

[6] Recall that with polysyllabic stems such as *uʔchi* the masdar has the same form as the primary stem, whereas monsyllabic stems reduplicate.

However, when we consider the form of the masdar formed from a monosyllabic root, a more differentiated picture emerges. Recall that monosyllabic stems obligatorily reduplicate in the masdar, as in (100)a. However, as (100)b shows, the presence of the benefactive V2 blocks the reduplication of a monosyllabic root in the masdar. This strongly suggests that the non-telic V2s are in fact part of the masdar itself, rather than following it, and with that part of the stem itself.

(100) a. *iŋ=aʔ* *bay-bay* *oʔ*
1SG=GEN make-RDP house
'the house I built'

b. *iŋ=aʔ* *bay kay* *oʔ*
1SG=GEN make BEN house
'the house that I built for [someone]'

On the other hand, telic V2s and the perfect were not accepted in this construction by speakers I consulted:

(101) **iŋ=aʔ* *bay=siʔ* *oʔ*
1SG=GEN make=PERF house
'the house I have built'

**iŋ=aʔ* *bay goʔɖ* *oʔ*
1SG=GEN make C:TEL house
'the house I built'

Thus, the V2s denoting telicity behave in this respect as a single unit along with the perfect, as (101) shows, and do not form part of the stem, while the non-telic V2s form part of the stem itself.

This analysis is supported by the fact that the semantic contribution of some non-telic V2s, although certainly not all, can be rather unpredictable with certain contentive morphemes, and not all V2s are compatible with all contentive morphemes. For example, the V2 *bay* is compatible with *aʔj* 'splash' and *gil* 'beat' with an "excessive" or "intensifying" meaning but also with *kayom* 'talk', with the resultant meaning 'deceive'. On the other hand, it is not compatible with *kamu* 'work' or *geʔb* 'burn', to name just two examples. Although this is not conclusive proof, such facts are at least compatible with an analysis of these units as forming a part of the stem. To my knowledge, no telic V2 shows such semantic or distributional restrictions.

If the non-telic V2s do in fact form a part of the stem together with the lexical root, we should then remove all non-telic V2s from the phrase-structure rules, as this process takes place in the lexicon, not in the syntax. This view is however also somewhat problematic, as V2s in many respects are phonological and syntactic words (cf. again example (47), where the floating clitic *=ga* intervenes between the "stem" and the V2 as well as the discussion in Peterson and Maas 2009), suggesting that they are not part of the stem.[7]

As the issue of the exact status of the non-telic V2s is as yet unresolved, these will not be dealt with further in this study and the telic V2s will be treated syntactically. However, the question as to whether they should be dealt with in the lexicon or in the syntax does not directly affect our argumentation, as both options are viable.[8] On the other hand, the telic V2s clearly do not form part of the stem and are syntactic atoms.

The preceding discussion forces us to again revise a few of the phrase-structure rules given above:

---

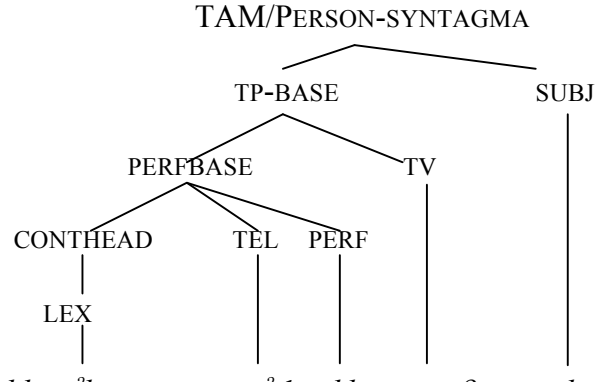
[7] Although (47) is an example of a telic V2, the non-telic V2s show similar behaviour with respect to the floating clitics.

[8] E.g., in Peterson (2006, chapter 8) the non-telic V2s were dealt with in the syntax, which caused no difficulties although it did require the assumption of an additional layer in the TAM/PERSON-SYNTAGMA.

(102) TAM/PERSON-SYNTAGMA → TP-BASE^{+} SUBJ
(103) TP-BASE → PERFBASE TV
(104) PERFBASE → CONTHEAD (TEL) (PERF)

I.e., the TAM/Person-syntagma consists of one or more TP-BASES and subject marking. The TP-BASE then consists of both a PERFBASE and TV (= TAM and basic voice), while PERFBASE consists of a CONTHEAD and, optionally, a telic V2 and marking for the perfect. The following presents an example of a TAM/Person-syntagma with both a telic V2 and the perfect marker.

(105)



'they had filled [it] up'

3.4.3 Minor types - The optative, present perfect and "Past II"

There are three further sub-types of the TAM/Person-syntagma from a purely structural perspective: Predicates marked for the optative, the "past II", and the present perfect. These three types differ structurally from other TAM/Person-syntagmas in that they are not marked for the active/middle distinction (TV) but can mark for the categories under PERFBASE such as TEL and the PERFect. The following presents one example of each type.

Optative

(106) *aniŋ=aʔ* *konsel* *kongher=ki* *khaɽiya=ki=yaʔ* *bair* *nãdani=te* *koŋ*
1PL.INCL=GEN girl boy=PL Kharia=PL=GEN old history=OBL know
guɽuʔ=may...
OPT=3PL
'Our girls and boys should know the old history of the Kharia …'[Kerketʈã 1990: ii]

Past II

(107) *ho=te* *kheti* *uslo?* *bes* *bes* *aw=khoʔ.* [MT, 1:76]
that=OBL(='there') field soil good REP QUAL=PST.II
'The fields and the soil there were good.'

Present perfect

(108) Q: *am=bar "Rock Garden"* *ɖe^{ʔ}b=siʔ=barʔ* A: *hã,* *ɖe^{ʔ}b=siʔɖ=iɲ.*
2=2HON "Rock Garden" ascend=PERF=2HON yes ascend=PERF=1SG
Q: 'Have you ever climbed "Rock Garden" (the name of a park on a hill in Ranchi)?'
A: 'Yes, I have climbed it.'

The optative and the past II are similar in that their respective markers do not show an active/middle distinction, although they do mark for the subject and appear to be compatible with TEL and PERF marking. This can easily be accounted for in the present format by assuming that these two markers are in complementary distribution with the other markers in tv and modifying this unit accordingly, as in (109)-(111), summarized in Table 4.

(109) TV = (TAM ∧ BV) ∨ OPT ∨ PTII
(110) TAM = (PAST ∨ PRESENT ∨ PROGRESSIVE ∨ IRREALIS)
(111) BV = ACTIVE ∨ MIDDLE

TAM ∧ BASIC VOICE	PAST ∧ ACTIVE	*=oʔ*
	PAST ∧ MIDDLE	*=ki*
	PRESENT ∧ ACTIVE	*=te*
	PRESENT ∧ MIDDLE	*=ta*
	PROGRESSIVE ∧ ACTIVE	*=te^{ʔ}j*
	PROGRESSIVE ∧ MIDDLE	*=ta^{ʔ}j*
	IRREALIS ∧ ACTIVE	*=e*
	IRREALIS ∧ MIDDLE	*=na*
OPT		*guɖuʔ*
PTII		*=khoʔ*

Table 4: TV (revised from Table 3)

The case is slightly different for the perfect marker. Like the optative and past II, the present perfect is unmarked for TAM/BASIC VOICE. Unlike these two markers, however, this is not an inherent trait of the perfect marker: While the present perfect with reference to a particular situation holding at the time of utterance is unmarked for TAM/BASIC VOICE, as in (108), the perfect marker may be followed by a TAM/BASIC voice marker such as the irrealis, past (active only), or the present (when referring to a habitual or iterative situation).

Past perfect

(112) *am=pe, iɲ* *gam=sikh=o^{ʔ}j* *ho=ghay=ga* *am=pe* *col=ki=pe.*
2=2PL 1SG say=PERF=A.PST.1SG that=way=FOC 2=2PL go=M.PST=2PL
'You, you went just as I had told [you].' [AK, 1:63]

Irrealis perfect

(113) *koroʔb=siʔ=na=pe.* *ber=jo* *i=jo* *aʔ=pe* *gam=e.*
silent=PERF=M.IRR=2PL who=ADD what=ADD NEG.MOD=2PL say=A.IRR
'Be quiet! Don't any of you say anything.' [Kerketʈā 1990: 2]

The lexical entry of the perfect will therefore have to be specified as optionally not co-occurring with TAM/BASIC voice marking. It will also have to be specified as being compatible only with the active when preceding past TV marking, never the middle voice.

4 Summary and outlook

With the few phrase-structure rules given in the preceding pages, the structure of "predicates" and their "complements" in Simdega Kharia can accurately be accounted for in a formalized analysis which makes use of neither nouns nor verbs. In fact, this analysis seems preferable to any analysis of Kharia in terms of these familiar lexical classes, since assuming their presence also forces one to assume a large number of zero-derivations and hidden verbs and even leads to incorrect forms, unless *ad-hoc* rules are formulated which apply only to these non-overt elements. This is not to say that it is impossible to analyze Kharia in terms of nouns, verbs and adjectives, only that the analysis presented here requires far fewer rules and makes far fewer theoretical assumptions. Following one of the basic tenets of descriptive linguistics, according to which categories are only to be assumed for a particular language if there is positive evidence for their existence, it seems most appropriate to favor an analysis of Kharia which does not assume these lexical classes, which would only be needed to make Kharia look more like other, more familiar languages.

There is one further, highly marginal type of event-narrating predicate which was not dealt with in the preceding pages and which differs slightly from the types discussed above – quotative predicates. The following is the only example I have come across so far but was accepted as grammatical by all speakers I consulted:

(114) *u* *buɽha=kiyar=te=ko* *bay* *jaʔb=siʔ.* *iɖib* *tunboʔ*
this old.man=DU=OBL=CNTR madness grab=PERF night daytime
"kersoŋ=e la! kersoŋe la!" *loʔ=na=kiyar.* [Kerketʈā 1991: 31]
marry=A.IRR VOC REP CNT=M.IRR=DU
'My parents have gone mad (= madness has grabbed the old man [and his wife = DU]). Day and night they'll keep on [say]ing "Marry! Marry!"'

What is different about this type of predicate is that what otherwise corresponds to the CONTHEAD is here marked for TAM/BASIC VOICE,[9] all of which should indicate that this is a TAM/Person-syntagma. If we accept this view – and everything seems to suggest we should – this means that we have here a TAM/Person-syntagma contained within yet another TAM/Person-syntagma.

[9] As well as PERSON marking, although this is negatively marked here, since the quotation is an imperative of the 2nd person, singular.

Although further study is required, examples such as (114) appear to be of a very different type than the TAM/Person-syntagmas discussed above. I suggest that the semantic base of (114), i.e., the quote, has been reanalyzed as a single syntactic unit or "syntactic word", following Di Sciullo and Williams (1987). Syntactic words in this sense are a class of objects that have syntactic form but otherwise display the general properties of X^0s. In other words, "The rules creating these objects essentially reanalyze a phrase as a word." (Di Sciullo and Williams 1987: 79). As Di Sciullo and Williams (1987: 79f.) show for the French expression *essui-glace* 'windshield wiper' (115) (gloss slightly adapted), which would appear to have the form "Verb-Object", a number of syntactic processes in French actually combine to show that *essui-glace* must be analyzed as a simple unit, such as the fact that an adjective modifies the entire unit (116), that it is not possible to modify the apparent verb *essui* with an adverb such as *bien* 'well' (117), and that no referential element can replace or modify *glace* 'glass' (118):

(115) *essui-glace*
wipe-glass
'windshield wiper'

(116) *un* *bon* *essui-glace*
indef good wipe-glass
'a good windshield wiper'

(117) **essui-bien glace*

(118) **essui-quelques-glace* **les-essuie*
wipe-some-glass them-wipe

I believe that the direct speech which is being quoted (114) is similar and could just as well have been an interjection such as "Ouch!" or "Hey!" – its internal form is unimportant here. As this is the only such example I have come across, such an analysis of course remains purely speculative and must await further verification but, in the absence of any counter-evidence, I will consider the semantic base in (114) to be an unanalyzable "syntactic word" in this sense. As this is restricted to quotative predicates, we need not reject the analysis developed in the previous pages but can simply view this type as an additional type of CONTHEAD which is functionally highly restricted.

In sum, the present analysis, although certainly requiring further study, is nevertheless the most economical (and theory-neutral) analysis to my knowledge which is capable of accounting for the Kharia data and which can easily be transferred to a number of different theoretical frameworks with more-or-less minor adaptations.

Finally, it is worth noting in conclusion that the present analysis by no means requires that the language in question be 100% flexible: Restrictions are bound to occur in any language, but whatever form they may take in a language such as Kharia and others like it, it would seem that such restrictions will unlikely be due to morphosyntactic properties of the respective lexemes but rather more likely to context, i.e., the feasability of certain semantic bases being found in certain functions, a possibility which should be thoroughly explored before resorting back to the more familiar – but at least equally problematic – lexical parts of speech.

Even if some lexemes should turn out to be restricted to either only the TAM/Person-syntagma or only the Case-syntagma, there is still no need to discard the present analysis entirely. Rather, these few lexemes will simply have to be marked in the lexicon as only being compatible with a particular functional head, without restructuring the entire lexicon to accomodate them. In other words, although Kharia would then be a language with nouns and verbs (or perhaps: with a noun or a verb), this search to prove the existence of such classes would tell us virtually nothing about how the rest of the language functions – including the vast majority of lexical morphemes – nor that what is of primary importance in Kharia is not the LEXICAL level but rather the SYNTACTIC level. Thus, by concentrating our discussion on whether or not a particular language can indeed be shown to have at least one noun and/or verb, or that these categories CAN be made to fit a particular language (with some ingenuity), we are probably overlooking quite a bit of other, potentially much more interesting information on languages of this type.

Abbreviations

A - active voice
ABL - ablative
ABS - absolutive
ADD - additive focus
ART - article
BEN - benefactive ("v2")
BV - basic voice
CAUS - causative
CL - classifier
CNTR - contrastive focus
Case-syntagma - the syntagma defined by the presence of case marking
C:TEL - culminatory telic
CLASS - classifier
DEF - definite
DU - dual
EXCESS - excessive ("v2")
FOC - restrictive focus
FUT - future
GEN - genitive
HON - honorific
HUM - human
INCL - inclusive (1/2, non-singular)
INF - infinitive
INST - instrumental
IPFV - imperfective
IRR - irrealis
M - middle voice
NEG - negative morpheme
NEG.MOD - non-indicative nega-tion
NHUM - non-human
OBL - oblique case
OPT - optative
PASS - passive
PERF - perfect
PL - plural
POSS - inalienable possession
PRST - presentative
PROG - progressive
PRS - present
PST - past
PT.II - "Past II"
QUAL - qualitative predication
RDP - reduplication
REP - (non-obligatory) repetition of an entire phonological word (intensity, distribution, etc.)
SG - singular
SUBJ - subject
TAM - tense, aspect, mood
TAM/Person-syntagma - the syntagma defined by the presence of tam / bv and person marking
TEL - telic v2
TV - tam and basic voice
V2 - unit denoting *Aktionsart*, which follows the lexical base of the TAM/Person-syntagma but precedes all other tam marking
VOC - vocative

References

Anderson, G. D. S. (2007). *The Munda Verb. Typological Perspectives* (Trends in Linguistics, Studies and Monographs, 174). Berlin and New York: Mouton de Gruyter.

Anderson, G. D. S. and Zide, N. H. (2002). 'Issues in Proto-Munda and Proto-Austroasiatic nominal derivation: the bimoraic constraint', in M. A. Macken (ed.), *Papers from the 10th annual meeting of the Southeast Asian Linguistics Society.* Tempe, AZ: Arizona State University, South East Asian Studies Program. (Monograph Series Press), 55-74.

Anderson, S. R. (2005). *Aspects of the Theory of Clitics* (Oxford Studies in Theoretical Linguistics, 11). Oxford: Oxford University Press.

Borer, H. (2005). *Structuring Sense.* Oxford: Oxford University Press.

Bresnan, J. (2001). *Lexical-Functional Syntax* (Blackwell Textbooks in Linguistics 16). Malden, MA - Oxford and Carlton, Victoria: Blackwell.

Broschart, J. (1997). 'Why Tongan does it differently: Categorial distinctions in a language without nouns and verbs', *Linguistic Typology* 1: 123-65.

Di Sciullo, A. and Williams, E. (1987). *On the Definition of Word.* Cambridge, MA: MIT.

Dixon, R. M. W. and Aikhenvald, A. Y. (2002). 'Word: a typological framework', in R. M. W. Dixon and A. Y. Aikhenvald (eds.), *Word: A cross-linguistic typology.* Cambridge: Cambridge University Press, 1-41.

Evans, N., and Osada, T. (2005). 'Mundari: the myth of a language without word classes', *Linguistic Typology* 9-3: 351-90.

Falk, Y.N. (2001). *Lexical-Functional Grammar. An Introduction to Parallel Constraint-Based Syntax.* Stanford, California: CSLI.

Hengeveld, K., and Rijkhoff, J. (2005). ‘Mundari as a flexible language', *Linguistic Typology* 9-3: 406-431.

Hoffmann, Revd. J., S. J. (1903). *Mundari Grammar.* Calcutta: Bengal Secretariat Press.

--- (1905/1909). *Mundari Grammar and Exercises. Parts I and II.* [Reprint: New Delhi: Gyan Publishing House, 2001].

Kerkeṭṭā, Kh. P. (1990). *jujhair ḍãṛ (khaṛiyā nāṭak)*. Ranchi: Janjātīya Bhāṣā Akādamī, Bihār Sarkār.

Malhotra, V. (1982). *The Structure of Kharia: A Study of Linguistic Typology and Language Change.* Ph.D. dissertation, Jawaharlal Nehru University (New Delhi).

Peterson, J. (2005). 'There's a grain of truth in every "myth", or, Why the discussion of lexical classes in Mundari isn't quite over yet.' *Linguistic Typology* 9-3: 391-441.

--- (2006). *Kharia. A South Munda Language. I: Grammatical Analysis. II: Kharia Texts. Glossed, Translated and Annotated. III: Kharia-English Lexicon.* Unpublished *Habilitationsschrift,* Universität Osnabrück.

--- (2007). 'Languages without nouns and verbs? An alternative to lexical classes in Kharia', in C. P. Masica (ed.), *Old and New Perspectives on South Asian Languages. Grammar and Semantics. Papers growing out of the Fifth International Conference on South Asian Linguistics (ICOSAL-5), held at Moscow, Russia in July 2003.* (MLBD Series in Linguistics, XVI), 274-303.

--- (2011). *A Grammar of Kharia. A South Munda Language.* (Brill's Studies in South and Southwest Asian Languages, 1). Leiden and Boston: Brill.

Peterson, J. and Maas, U. (2009). 'Reduplication in Kharia : the masdar as a phonologically motivated grammatical category', *Morphology* 19-2: 207-237.
Pinnow, H.-J. (1965). *Kharia-Texte (Prosa und Poesie).* Wiesbaden: Harrassowitz.
Sadock, J. M. (1991). *Autolexical Syntax. A Theory of Parallel Grammatical Representations* (Studies in Contemporary Linguistics). Chicago and London: University of Chicago Press.
Simpson, J. (1991). *Warlpiri Morpho-Syntax. A Lexicalist Approach* (Studies in Natural Language and Linguistic Theory 23). Dordrecht: Kluwer.

Proper names, predicates, and the parts-of-speech system of Santali

Felix Rau*

1 Introduction

The present paper gives preliminary observations on the syntax and semantics of proper names in the North Munda language Santali. Munda languages, such as Mundari, Santali and Kharia, have been reported to exhibit a large degree of flexibility in their parts-of-speech systems. The focus of this paper is on proper names in predicate position as an extreme case of flexibility, because proper names are generally thought to be referring expressions that cannot easily be given a predicative reading. Given these characteristics, the combination of predication and proper names provides a good test case for the flexibility of lexemes and generality of this characteristic.

I will argue for an account of the parts-of-speech system of Santali emphasizing the flexibility of categories. Contrary to claims that this flexibility is restricted and displays semantic irregularities, I will try to show that in Santali flexibility is indeed regular and a general characteristic of the language and that it is carried to extremes which defy derivational explanations. By focusing on a small phenomenon – instead of analysing the whole parts-of-speech system of Santali – I hope to contribute to the discussion on the parts-of-speech systems in Munda languages that has been revived in recent years (cf. Evans and Osada 2005; Peterson 2005; Hengeveld and Rijkhoff 2005), broaden the empirical base and thus enhance our understanding of Santali and hopefully the Munda languages in general.

The claims I make are about Santali only, but I will often contrast my findings with the statements made by Evans and Osada (2005) about Mundari. I believe that what is said here could have some bearing on the analysis of Mundari and probably other Munda languages. The general direction of my analysis has been argued for by Peterson (2005, this volume). It is also relevant to note that not all Munda languages seem to display such a flexibility in their parts-of-speech systems and most South Munda languages such as Sora, Gorum and Gutob – with which I am much more familiar – are quite different in this domain and display much less flexibility than the North Munda languages and Kharia.

The statements made in this paper are based mainly on data from the texts of Bodding (1926; 1927; 1929b). In the grammatical analysis of Santali I will generally follow Neukom and make it explicit if I deviate from his analysis.[1] I also use a slightly modified version of his spelling conventions for the examples taken from Bodding.[2]

---

* I would like to thank Eva van Lier, Jan Rijkhoff, Arlo Griffiths and the two anonymous reviewers for their very helpful comments.

[1] I use the same glosses as are used by Neukom (2001): ACT active, AN animate, APPL applicative, COP copula, CPL completive, D/DU dual, DAT dative, E/EXCL exclusive, FOC focus, GEN genitive, I inclusive, INAN inanimate, IND indicative, INST instrumental, IPFV imperfective, MID middle, MOD modal, NML nominalizer, O object, P/PL plural, POSS possessive/possessor, PROH prohibitive, PST past, S singular, S subject, TOP topic. Several of the sentences I have taken from Bodding (1926; 1927; 1929b) are also used in Neukom (2001). In these cases, I have only cited the original source.

[2] I use the IPA characters ɖ and ʈ instead of the indological style ḍ and ṭ to write the retroflex phonemes.

2 Previous accounts

The parts-of-speech systems of Munda languages (especially the North Munda languages Santali and Mundari) have been subject to several analyses. There are two main lines of argument: the first focuses on the variety of syntactic contexts a given lexeme can occur in and emphasizes the general validity of this property. This approach is taken by majority of scholars beginning with Hoffmann for Mundari and Bodding for Santali to Hengeveld (1992) and Hengeveld and Rijkhoff (2005) for Mundari, Neukom (2001) for Santali, as well as Peterson (2005) for Kharia. The second approach, mainly represented by Evans and Osada (2005), focuses on inequalities in the distribution of lexemes and argues for a more ordinary parts-of-speech system with nouns, verbs and adjectives. This approach is in fact also taken by most of the lexicographic work on North-Munda languages (e.g. Bodding 1929-1936).

2.1 The flexibility analysis

The traditional analysis regards the North Munda languages Mundari and Santali as clear cases of flexible languages with only one major word class (cf. Hengeveld et al., 2004, for a comprehensive typology of parts-of-speech systems), in which every lexeme can occur in every syntactic postion.[3] The traditional approach is perhaps best outlined by Hoffmann (1903):

> Instead, then, of Parts of Speech with well-defined functions and a precise but rich denotative and connotative power, we meet in Mundari with words of great functional elasticity, and therefore of a vague signifying power – words which, whilst *denoting* living beings, actions, qualities, and relations, do generally not by themselves *connote* the manner in which the mind conceives the things signified. That *connotation* is generally left to the context of the proposition or the circumstances under which it is uttered; [...] (Hoffmann 1903: xx–xxi; original emphasis)

An opposing analysis is that taken by Evans and Osada (2005), on which I will concentrate in the following section. This analysis is unique in its data coverage of a specific Munda language (i.e. Mundari) and rejects the flexibity analysis.

2.2 The analysis of Evans and Osada (2005)

Evans and Osada (2005) propose an analysis of the Mundari parts-of-speech system which claims "that Mundari clearly distinguishes nouns from verbs, though (like English, Chinese, and many other languages) it has widespread zero conversion" (Evans and Osada 2005: 384). They posit three criteria as requisites for a flexibility analysis: equivalent combinatorics of all lexemes, compositionality of semantics in all functions and bidirectionality of flexibility (Evans and Osada 2005: 366). In their view, what seems to be flexibility of categories in Mundari does not meet their criteria and is in fact

[3] Only Bhat (1997) regards Munda languages as omnipredicative and thus as rigid languages with only one word class.

conversion or zero derivation. Unfortunately they do not specify what exactly they understand by conversion, but as a lexical process, it should apply to lexemes only and its semantic effect should be lexeme-specific and not regular. This last claim is stated explicitly in Evans and Osada (2005: 374). Another property of a lexical derivational process such as conversion would be that there is no requirement for it to apply to all lexemes. In fact, Evans and Osada (2005: 384) claim that the process does not extend over the whole lexicon and is thus not general, but they estimate 50% of the lexicon to be subject to the process of conversion.

Peterson (2005) as well as Hengeveld and Rijkhoff (2005) have argued convincingly against this analysis, although they differ in what they accept as requisites for flexibility. Most importantly, Hengeveld and Rijkhoff (2005) reject the requirement of semantic compositionality. However, the claims made by Evans and Osada (2005) that flexibility is not a general property of the lexemes and that many alleged examples of flexibility show non-compositional lexeme-specific idiosyncrasies still stands.

3 Arguments and predicates in Santali

In this section, I will outline two aspects of the grammar of Santali relevant to the present purpose, before going on to the actual focus of this paper, i.e. predication and proper names. After introducing the main predicate position of a Santali sentence and some of its syntactic and morphological characteristics, I will sketch the behaviour of arguments and adjuncts in Santali syntax. For a more comprehensive account of the grammar of Santali see Neukom (2001), Ghosh (2008) and Bodding (1929a).

3.1 Main predicate position

A sentence in Santali can contain several predicates, but there may only be one main predicate. This main predicate has a unique morphological device that identifies it: the so-called indicative suffix *-a* (cf. Neukom 2001: 145).[4] Other verbal morphology, such as TAM suffixes and person markers, is not confined to the main predicate position, but may occur on all predicates.

The main predicate position in Santali syntax can be best delimited by two morpho-syntactic clues: the above-mentioned indicative suffix *-a,* which occurs at the right edge of the predicate position, and the subject clitic, which according to Neukom (2001: 113) is "normally attached to the word that immediately precedes the verb", but can also follow the indicative suffix *-a* in some situations (ibid.). The wording "immediately precedes the verb" is somewhat unfortunate, since there are several examples in which the subject clitic does not precede the lexeme one would like to call a verb, but the predicate-constituting syntagma that is predicated, as in example (1). In this sentence the clitic precedes the whole ArgP that forms the main predicate in this sentence.

[4] This is only true for some sentence types. Most non-asserted sentences such as imperatives and presupposed clauses lack this marker. Sentences of this kind will not be discussed in this paper. However, subordinated clauses, which do not take this marker either, will be briefly discussed together with arguments.

(1) *ape-pe* [*mɔ̃ɽɛ̃* *hɔɽ-a*]
you(PL)-2pS five people-IND
'You are five (people).' (Bodding 1929b: 350)

Sentences such as (1) show how the main predicate is morphosyntactically delimited. The left boundary is marked by the subject clitic, which attaches to the material preceding the predicate, while the right boundary is situated directly after the indicative marker. Hence, the structure of the main predicate in a sentence can be represented as in (2).

(2) NON-PREDICATED PART-SUBJ [MAIN.PREDICATE -IND]

The subject clitic may also stand in a different position, directly after the indicative marker *-a*. This is the case in at least two circumstances. On the one hand, being an enclitic, the subject clitic obviously requires some material to its left to which it can attach. If no material precedes the predicate, the subject clitic is attached to the right of the predicate. A single-lexeme sentence as in example (3) is a typical example for this behaviour.

(3) *dal-et'-kan-a-e*
strike-IPFV:ACT-IPFV-IND-3sS
'He is striking.' (Neukom 2001: 64)

Another, more complicated, case – which at closer inspection may nevertheless turn out to be quite similar to the previous one – is the rare situation in which more than one lexeme constitutes the sentence, while the subject clitic, contrary to expectation, follows the indicative marker as in example (4). In some cases this might indicate that the sentence should be regarded as thetic, which again would result in a situation where there is no material preceding the main predicate. Yet in other cases this seems unlikely and the motivation for the placement of the subject clitic in sentence-final position remains unclear.

(4) *ale-dɔ* *lelha* *bhucuŋ* *koŋka* *Bhũiə* *kan-a-le*
we(PL.EXCL)-TOP stupid ignorant foolish Bhuya COP-IND-1peS
'We are foolish, stupid, witless Bhuyas.' (Bodding 1929b: 350; cf. also Neukom 2001: 114)

Virtually every lexeme can form a main predicate in Santali. Besides event-denoting lexemes such as *dal* 'to strike' in example (3) above, entity- and property-denoting lexemes such as *raj* 'king' and *maraŋ* 'big' in example (5) and (6) respectively can also be used as predicates. The main predicate does not even have to be an actual lexeme, as is the case with the onomatopoetic *ãã* in (7).

(5) *adɔ-e* *raj-en-a*
then-3sS king-PST:MID-IND
'So he became king.' (Bodding 1929b: 8)

(6) *bəjun-dɔ-e* *maraŋ-a*
Bajun-TOP-3SS big-IND
'Bajun is big (senior)' (Bodding 1927: 2 [translation F.R.])

(7) *bar* *pɛ dhao-e* *ãã-y-en-a*
two three.time-3SS groan-y-PST:MID-IND
'It (i.e. the buffalo) groaned two or three times.' (Neukom 2001: 15)

Furthermore, the main predicate can consist of non-lexical units such as phrases. The postpositional phrase *kombɽo tuluj* 'with thieves' in (8) is an example for this kind of main predicate. The point that phrasal and other complex units can function as predicates in Munda languages has been made before (Peterson 2005: 396f.) and it is in itself a very good argument against analyzing the flexibility a lexical derivational process.

(8) *alo-m* *kombɽo* *tuluj-ok'-a*
PROH-2SS thief with-MID-IND
'Don't keep company with thieves.' (Neukom 2001: 15)

The common property of all these predicates is that they always get a predicative interpretation and are never referential. Also, main predicates are always asserted and never presupposed. These semantic characteristics of the main predicate position are crucial for the present argument.

3.2 Argument position

The argument positions of a sentence in Santali are occupied by argument phrases (cf. Neukom 2000: 18), such as the definite argument phrase in (9). Argument phrases (ArgP) consist at least of a head element – such as *koɽa gidrə* 'boy child' – which may be preceded by modifiers, e.g. the reduplicated *huɖiɲ* 'small' in example (10), and adnominal demonstratives, such as *uni* in both of the sentences below.

(9) *uni* *koɽa* *gidrə-dɔ*
that(AN) boy child-TOP
'that boy (child)' (Bodding 1929b: 86)

(10) *uni* *huɖiɲ* *huɖiɲ* *gidrə-dɔ*
that(AN) small small child-TOP
'that very small child' (Bodding 1929b: 84)

Predicates and whole clauses can also be part of an argument phrase. In example (11) a complete clause, consisting of a compound verb with tense affix and subject clitic, occupies the modifier position in the ArgP. The subject clitic is attached to the demonstrative that is part of the ArgP and precedes the clause.

(11) *uni-y*[-e *bujhəu-ɲɔ̃k′-ket′*] *hɔɽ*
that(AN)-y-3SS understand-little-PST:ACT person
'The man who understands little' (Bodding 1926: 18)

ArgPs can be marked for case. This marking is done by suffixes such as *-ʈhɛn* in example (12). These suffixes are also attached to subordinated clauses as in (13), where the same suffix *-ʈhɛn* functions as a subordinator. These clauses are syntactically and morphologically complete, except that they cannot contain the indicative marker *-a*. Nevertheless they display case marking and behave syntactically like any other case-marked ArgP.

(12) *algel hɔɽ-ʈhɛn-dɔ alo-pe ləi-a*
outside person-DAT-TOP PROH-2PS tell-IND
'Don't tell it to outsiders' (Neukom 2001: 24)

(13) *gapa-dɔ am-ge si-ok′-ʈhɛn ɖaŋgra-dɔ laga-əgu-kin-me*
tomorrow-TOP you(s)-FOC plough-MID-DAT bullock-TOP drive-bring-3DO-2SS
'Tomorrow you shall drive the bullocks to where I am ploughing'
(Bodding 1926: 100)

The structure and morphology of the ArgP is independent of the category of the elements that form its constituents. There seem to be no general restrictions on what category an element must belong to so as to function as a head or modifier in an ArgP.[5]

Thus, in summary, the relevant facts about arguments and predicates in Santali are the following: ArgPs in argument position are referential or quantifying expressions with a specific structure and morphology. Predicates can be part and even head of a ArgP, but lack the indicative marker *-a* in this position. Semantically, predicates or clauses that take the form of an ArgP are not asserted, but presupposed. On the other hand, nominals as well as complete ArgPs and PPs can be placed in the main predicate position, just like event- or property-denoting lexemes. Any material that forms the main predicate is non-referential, regardless of its lexical semantics.

4 Nominal sentences

In light of the main concern of this paper, namely the semantics and syntactic properties of the predicate position in Santali, I will concentrate in the following on a special case of predication, the so-called nominal sentence. Nominal sentences are a group of sentence types that are defined by a common syntactic property: the predicating element is formally not a verb, but an ArgP (or an AP or PP). In some languages, such sentences

[5] The only exception seems to be bi- or trivalent lexemes, which do not occur in head position without having their argument positions satisfied by either detransitivation or the use of arguments. Given the connection between the existence of transitive lexemes and the presence of a noun-verb distinction made by Rijkhoff (2003), this fact might be evidence against the idea that the Santali parts-of-speech system is entirely flexible. A detailed study of transitivity in Santali is still a desideratum, hence nothing more can be said about this topic here.

involve the use of a special copular verb, while in other languages the predicate ArgP and the subject ArgP are juxtaposed without an element of this kind. Regardless of their language-specific form, nominal sentences can be divided into different types according to their meaning.

There are different accounts of the types of nominal sentences, starting with Higgins (1979), and no consensus has been reached on their number or what their characteristics are. I will confine my discussion of nominal sentences to a very simplistic typology of three types (following Mikkelsen 2005): predicational, specificational, and equative. Mikkelsen (2005: 58) gives clear exemplary sentences for these three types. Example (14) is unambiguously a predicational sentence, example (15) is unambiguously equative, and example (16) is clearly, although not unambiguously, a specificational sentence.

(14) *The winner is Republican*. (Mikkelsen 2005: 58, example 4.18)
(15) *He is McGovern*. (Mikkelsen 2005: 58, example 4.20)
(16) *The winner is Nixon*. (Mikkelsen 2005: 58, example 4.19)

Predicational sentences such as (14) ascribe a property, denoted by the complement, to the subject (Declerck 1988; Mikkelsen 2005) and are thus closer in their semantic structure to verbal sentences than the other two types. Equative sentences actually predicate the fact that the referents of the two nominals involved are the same entity. Sentences such as (15) are exceptional in that they have a referential nominal – here the proper name *McGovern* – as their complement. Finally, specificational sentences are special in two respects; they have a referential nominal in the complement – the proper name *Nixon* in example (16) – and a property-denoting (or predicational)[6] expression in the subject position. However, sentences like (16) are ambiguous because definite ArgPs such as *the winner* can also be interpreted as referential, depending on the context; this would render the sentence an equative one. Specificational sentences are highly restricted by discourse and are frequently ambiguous, even in a concrete context, between specificational and equative readings.

The three different types of sentences can thus be characterized by the distribution of referential and non-referential nominals in subject and complement position. Table 1 lists the different types and the referential status of their subjects and complements (after Mikkelsen 2005: 50).

	Subject	Complement
Predicational	referential	non-referential
Specificational	non-referential	referential
Equative	referential	referential

Table 1. Subjects and Complements

I will disregard specificational sentences in Santali for the time being, because my data are insufficient to present a substantiated discussion of this problematic sentence type.

[6] I follow Mikkelsen (2005: 51) in using the terms *property-denoting* and *predicative* interchangeably. Formally, this requires viewing properties as functions from individuals to functions from worlds to truth values.

The semantic difference between the predicates in predicational and equative sentences is sufficient for the purpose of this paper and corresponds with significant differences in syntactic structure.

In Santali, nominal sentences are composed of a subject, a complement and often an element *kan*, which has been analyzed as a copula. The difference between predicational and equative sentences is minimal at first sight. Predicational sentences have the subject clitic attached to the subject, for example *-e* in (17), while the clitic occurs at the end of the whole sentence in equational sentences. Example (18) shows the latter variant.[7]

(17) *mit'-dɔ-e* *bhut* *kan-a*
one-TOP-3SS ghost COP-IND
'One was a ghost' (Bodding 1926: 8)

(18) *nui* *ma* *iɲ-ren* *hɔɽ* *kan-e*
this(AN) MOD 1S-GEN:AN person COP-3SS
'This is my wife' (Bodding 1926: 6)

The placement of the subject clitic is relevant for the delimitation of the main predicate. Thus on a closer inspection of the structure of the predicational sentence (17), we see that the nominal in the complement is situated inside the main predicate position. Yet at first sight, it does not appear to function as the predicate on its own, but is accompanied by the copula. This situation seems to support a claim made by Evans and Osada that nominals cannot constitute a predicate on their own. To support their claim, they present nominals in predicative position (Evans and Osada 2005: 371). As their examples (24b) and (25b) – here given as (19) and (20), respectively – seem to denote more complex events than one would like to see arising from the lexical semantics of the nominal, Evans and Osada take these sentences as evidence that a more complex derivational process must be assumed to account for the semantics of these sentences.

(19) *dasi-aka-n-a=ko*
servant[8]-INIT.PROG-INTR[9]-INDIC=3PL.S
'(They) are working as servants' (Evans and Osada 2005: 369, example (24b))

(20) *soma=eq* *baRae-aka-n-a*
Soma=3SG.S baRae-INIT.PROG-INTR-INDIC
'Soma has become a baRae.' (Evans and Osada 2005: 370, example (25b))

Evans and Osada (2005: 370) rightly focus on the compositionality of the semantics of these sentences and state: "First, one would still need to find an aspect allowing *mastaR*, *baRae*, *baa*, etc. to be used in the exactly composed meaning 'be a teacher', 'be a

[7] The example lacks the indicative suffix *-a*, because the modal particle *ma* is present (cf. Neukom (2001: 162) where sentence (18) also appears). This particle is used when the speaker assumes the statement to be known, but it has no influence on the placement of the subject clitic.
[8] The lexeme *dasi* is glossed 'serve' by Evans and Osada (2005: 369). Since they themselves use the gloss 'servant' in their article in example (24a) (ibid.) and it does mean 'servant' in referential use and 'work as at servant' and not 'to serve' in predicational use here in this example I changed their gloss for clarity.
[9] The gloss INTR for *-n* is missing in Evans and Osada (2005: 369).

blacksmith', 'be a servant', etc." (Evans and Osada 2005: 371). Although they are right to expound the problems of these sentences and mention the potential influence of the TAM categories, the examples given by them contain a suffix *-aka* glossed *initiated progressive*, whose contribution to the semantics is not discussed. With this problem in mind, let me come back to our Santali example (17) and the copula *kan*.

The element *kan* in Santali can be interpreted as a copula or an imperfective suffix (cf. the remarks in Neukom, 2001: 17). There are no morphological properties that distinguish the two in constructions such as example (17). The imperfective suffix *-kan* is part of the TAM morphology and is used to express ongoing and habitual actions as well as durative aspect. However, if -kan can be regarded as part of the morphological paradigm, it might be interesting to study the predicative use of alleged nominals in other forms of this paradigm.[10] Neukom (2001: 110) presents the lexeme *ʈuər* 'orphan' in three TAM forms. The first form is the zero-marked stative[11] in example (21), which exhibits a genuine predicational meaning. The second occurrence of *ʈuər* 'orphan' is part of a complex predicate (Neukom 2001: 142) and carries a completive past active suffix; it has basically causative semantics. The last example (23) displays middle voice marking, which yields inchoative semantics.

(21) *alaŋ-dɔ-laŋ* *ʈuər-ge-a*
we(di)-TOP-1DIS orphan-FOC-IND
'We are orphans' (Neukom 2001: 110, example (18a) originally from Bodding)

(22) *huɖjɲ* *gidrə-i* *ʈuər-oʈo-e-a*
small child-3SS orphan-leave-CPL:PST:ACT-3SO-IND
'She left a child motherless.' (Neukom 2001: 110, example (18b))

(23) *khange* *uni* *gidrə-dɔ-e* *ʈuər-en-a*
then that(AN) child-TOP-3SS orphan-PST:MID-IND
'Then the child became an orphan.' (Neukom 2001: 110, example (18c) originally from Bodding)

These examples show that lexemes such as *ʈuər* 'orphan' show a regular behaviour in predicate position, similar to that of event-denoting lexemes and most closely resembling that of state- or property-denoting lexemes, which all display a "to be x" meaning in stative usage, causative semantics such as "to make x" in active voice and an inchoative "to become x" reading in middle voice (cf. Neukom 2001: 109). The semantic differences between sentences with the imperfective suffix (or copular) *-kan* as example (17) and sentences such as (21), which lack a TAM suffix, seem to be minimal and probably restricted to aspectual subtleties. The instances of such structures in the texts of Bodding (1926; 1927; 1929b) are not very telling with respect to the semantic differences between the nominal sentences in active non-past and imperfective form.

---

[10] There is, however, a past tense copula *tah ẽkan*, which behaves more like an independent lexeme than *kan* (cf. Neukom 2001: 171ff).
[11] Zero-marked forms are normally analyzed as active non-past in Santali (cf. Neukom 2001: 62).

These structures need to be tested with speakers using sophisticated tests for aspect and aktionsart.

The data presented in this section demonstrate how general the predicative semantics of the main predicate position are. The referential complement of an equative sentence is realized as an argument, while the non-referential complement of a predicational sentence is realized as the main predicate of the sentence. Furthermore, the predicative usage of these supposed nominals seems to be remarkably regular in its semantics and closely parallels the semantics of event-denoting lexemes. This evidence clearly favours an analysis that assumes the flexibility of the lexemes involved and explains the different semantics via the syntactic positions in which they occur. Nevertheless, one could still argue for the existence of two different words such as a nominal orphan$_N$ and verbal orphan$_V$, where the latter is derived from the nominal *ʈuər* 'orphan'. To show that this lexical line of argumentation runs into problems, I will focus in the next section on lexemes that have no affinity to a property interpretation, namely proper names.

5 Predicate position and proper names

Proper names are a special kind of nominal expressions. They are used to name and to refer to individuals and are thought to be directly referring expressions and inherently definite. Like indexicals, proper names depend highly on the context to determine their referent. Usually, they constitute a lexical subclass of a category *noun* or a small independent category among the nominals (cf. Anderson 2007; Van Langendonck 2007), but are generally neglected in the discussion of lexical categories and parts-of-speech systems. This is probably due to their marginal status among the *nominals* and their inherent tendency to occur in referential usage only.

Their dominantly referential function and their context-dependence makes poper names a good test case for flexibility of the part-of-speech system of Santali. It is difficult to imagine a language with a fully productive mechanism to derive a verbal lexeme with predictable semantics from a proper name, given the latter's characteristics as sketched above.

The precise semantics of proper names are disputed, but there are two main lines of reasoning, which are referred to respectively as rigid designator theory and definite description theory. The rigid designator theory as advocated by Kripke (1972) and many others assumes that proper names are directly referring rigid designators (or indexicals). In this analysis, proper names cannot occur as predicates, since they are directly referring. Proper names that do seem to occur as predicates have to be explained as either not actually being predicated or not being used in the predicate, but merely mentioned. This line of reasoning is not very enlightening for my present purpose, so I will follow the other line of argumentation, stemming from Frege (1893) and Kneale (1962), which sees them as definite descriptions of some kind. By equating them semantically to definite descriptions, proper names are not referential by themselves in this theory, but quantificational (at least from a Russelian perspective). This makes proper names compatible with predication, although they can be – and most frequently are – used to refer.

The definite description theory has been defended by Geurts (1997) and Matushansky (2005, 2009) under the name *quotation theory*, because it assumes that the proper name is quoted in its semantics. Under this theory, a proper name *N* means *the individual N* following Geurts (1997), while Matushansky (2005, 2009), by reference to

Recanati (1993), adds a naming convention which results in something like *x is referent of [N] by virtue of the naming convention R*. For Bach (2002), in his nominal description theory, a proper name *N* means *the bearer of 'N'*, but in contrast to the quotational theory, Bach (2002: 76) views the meaning of proper names not as quotational but reflexive. This avoids some of the complications that come with the notion of quotation. Since there is no consensus on the exact semantics of a proper name within the definite description theory, I will for now assume a meaning along the lines of *N = the individual named N*. Thus I would like to keep the naming relation explicit in the semantics of proper names, but do not commit myself to a precise formal analysis or the quotational or reflexive character of its semantics.

Even under a definite description analysis, non-referential use of proper names is a marginal case, but at least a possibility, and one would not expect proper names to have the same distribution as state- or property-denoting lexemes.
In nominal sentences proper names are expected to occur mainly in the complement position of equative sentences, as *McGovern* in example (15), and in the subject position of predicative sentences, as in example (6), here repeated as (24).

(24) *Bəjun-dɔ-e* *maraŋ-a*
Bajun-TOP-3SS big-IND
'Bajun is big (senior)' (Bodding 1927: 2 [translation F.R.])

In the light of the discussion of the semantics of the predicate position in the previous section, the function and semantics of proper names are unlikely to be compatible with the predicate position. The claim made by Evans and Osada (2005: 371) "that proper names, such as Ranci 'Ranchi' [a toponym (F.R.)], are unavailable for predicate use" does therefore not seem to be a serious objection against a flexibility analysis of Munda languages. After all, proper names are normally used referentially, and the predicate position in Santali forces a property-denoting and thus non-referential reading.
The only type of sentences in which referential expressions are expected to occur in complement position are equative sentences. Since the complements are referential, they should not occur in pedicate position in Santali. And in fact, equative sentences as example (25) have the structure SUBJ COMP [COP]$_{PRED}$ with the proper name in the complement argument position and only a copula in predicate position.[12]

(25) (context: Two jackals who feature in the story turn out to be a god in disguise)
unkin-dɔ *Cando-ge*[13][*-kin* *tahɛ̃kan-a*]
that(AN):d-TOP Chando-FOC-3DS COP:PST-IND
'They were Chando himself.' (Bodding 1926: 160)

There are also examples where the copula is missing, as in (26), but the crucial point is that the proper name has still the form of a complement and does not carry the

[12] Specificational sentences also have a referential complement, but as noted above it is unknown how this sentence type is formed in Santali. As stated above, I exclude them from the discussion here.
[13] The focus marker *ge* is not part of the equative structure, but due to the contrastive context in which this sentence stands.

indicative suffix, nor is it preceded by the subject clitic. The predicate position of these sentences has to be analyzed as empty.

(26) *iɲ-dɔ Sitəʈi jugi*
I-TOP Sitari yogi
'I am yogi Sitari.'[14] (Bodding 1929b: 24)

There are, however, some contexts in which proper names are used predicatively (cf. Neukom 2001: 14) and thus non-referentially. The act of naming is probably the most prominent example of an arguably predicative use. Interestingly, this is the only usage in which proper names occur in the predicate position in Bodding (1926; 1927; 1929b). In a naming construction the proper name occurs as an applicative[15] in the predicate position. In example (27) the ArgP headed by *ɲutum* 'name' is the subject of the sentence and the ArgP headed by *hɔpɔn* 'son' functions as the object. The proper name *Turtə* functions as a transitive predicate that is combined with the two arguments. The sentence structure can be mimicked by the "literal" translation 'The boy's name is Turta-ing the woman's son'. This is in fact given in Neukom as an alternative translation.[16]

(27) *uni buɖhi-ren hɔpɔn-tɛt' koʈa-w-ak' ɲutum-dɔ*
that(AN) old.woman-GEN:AN son-3POSS boy-W-NML:INAN name-TOP
Turtə-w-a-e-a
T.-W-APPL-3SO-IND
'The old woman's son's name was Turta.' (Bodding 1926: 110)

The argument *ɲutum* 'name' is not necessary for the naming reading. Sentence (28) is a case in point (context: in a village there lived two people, mother and son). The structure of this sentence could be mimicked – along the lines of Neukom's "literal" translation – as 'Her son was Anua-ed' .

(28) *hɔpɔn-tɛt'-dɔ Ənuə-a-e-a*
son-3POSS-TOP Anua-APPL-3SO-IND
'The name of the son was Anua.' (Bodding 1926: 98)

The semantics of the proper names in these naming sentences can reasonably be

---

[14] Bodding translates this sentence as 'My name is Sitari jugi.' This is probably a more idiomatic translation in the context of this story, but the sentence structure and a comparison to other sentences strongly favour a translation as given above.
[15] For a description of the applicative in Santali, see Neukom (2001: 120).
[16] There is an alternative naming construction in Santali, which involves a proper name in complement position (like *Bitnə* in the following sentence) and the lexeme *ɲutum* meaning 'name' in predicate position:

ona-te-ge uni-dɔ Bitnə-ko ɲutum-kad-e-a
that(INAN)-INST-FOC that(AN)-FOC Span-3pS name-CPL-3SO-IND
'Therefore they called (named [F.R.]) him Span.' (Bodding 1927: 150)

analyzed as compositional, if we assume semantics along the lines of a definite description theory, and if the naming relation is regarded as a part of its semantics. In sentence (28), the description individual named Anua is made into a transitive property by means of the applicative suffix -a and ascribed to *hɔpɔn* 'son' by some unnamed agent (hence the passive in the translation), which could be paraphrased as S is made the individual named Anua (by X). Since being made the individual named N is actually being the object in the act of naming, the naming semantics seem to fall out naturally if we assume a meaning *the individual named N* for proper names. Since there is independent evidence for this assumption (cf. Matushansky 2005; 2009), these sentences and their semantics are a strong case in favour of the regular semantic properties of the predicate position and the flexibility of lexemes in Santali, as the sentence meaning can be derived compositionally from the lexical semantics of the proper name, the applicative, and the predicative semantics of the main predicative position.

There are also complex syntactic units that contain proper names and occur in predicate position. Sentence (29) contains two coordinated proper names in its main predicate position and could be paraphrased as 'They were Kara-and-Guja-ed.' Although the semantics of this sentence are complicated by the necessarily distributive reading of the predicate, it is essentially identical in structure with the examples given above. It is clear that an example such as (29) cannot possibly be analyzed on a lexical level.

(29) *unkin-dɔ* *Kaɽa ar* *Guja-w-a-kin-a*
that(AN):d-TOP K. and G.-w-APPL-3dO-IND
'Their names were Kara and Guja.' (Bodding 1927: 162)

In light of the remarkable ability of proper names to be used as transitive predicates in naming constructions, it would be interesting to see how these lexemes behave synactically and semantically as intransitive predicates. Whereas there is no instance of such a usage in the corpus of Bodding (1926; 1927; 1929b), Neukom (2001: 14) fortunately provides an (elicited) example:

(30) *uni-dɔ-e* *Nɔndɔ-a*
that(AN)-TOP-3sS Nanda-IND
'He acts as Nanda.' (Neukom 2001: 14)

In this sentence the bare proper name *Nɔndɔ*, without an applicative suffix, constitutes the predicate of the sentence. Such a form, with only the subject clitic and the indicative suffix, is the neutral non-past active form, already seen with the nominal *ʈuər* 'orphan' in example (21). Since example (30) is elicited and thus without a specific context, its semantics can only by deduced from the English translation, "to act as x", and it may be interpreted as instantiating the relevant properties for being x for some extent of time. This meaning is more or less in accordance with the semantics of a neutral non-past active, which is used to express habits and states (cf. Neukom 2001: 65).

The examples given throughout this section show that Santali allows proper names to be used as predicative expressions and that they occur in these cases in the main

predicate position. The semantics of these constructions are reasonably regular. The non-past active form ascribes the (temporal) property of *being the individual named N* to the subject, while the applicative form denotes that *x is caused to be the individual named N*, which translates into *being called N*. The account of the semantics of these cases is still very preliminary, and especially the exact contribution of the verbal morphology needs to be investigated; however, I think the general pattern is clear enough.

In this context, two examples from Kharia given by Peterson (2005: 395), reproduced here as (31) and (32), are interesting. In these examples the toponym *a?ghrom* is used in two different TAM forms as a predicate. The translations of these examples show some differences from the semantics of the Santali examples. In both cases the toponym is not marked by an applicative, but both display naming semantics. The middle voice past form of example (31) is inchoative in meaning while the active past example denotes the causative act of naming. It would be interesting to see whether these differences can be explained by the different semantics of the TAM categories in the two languages.

(31) *a?ghrom=ki*
Aghrom=MID.PST
'became/came to be called "Aghrom".' (Peterson 2005: 395)

(32) *a?ghrom=o?*
Aghrom=ACT.PST
'S/he made/named [the town] "Aghrom".' (Peterson 2005: 395)

Despite the limitation of the data, I hope to have shown convincingly that proper names can occur as predicates in the main predicate position in Santali and that their behaviour is parallel to that of other lexemes with descriptional semantics and even state- and event-denoting lexemes.

6 Conclusion

With this paper I intend to contribute to the analysis of the parts-of-speech system of Santali, and in focusing on the predicate position, I hope to have shown that lexemes in this language are in fact flexible to a remarkable degree. While the use of seemingly nominal lexemes as predicates in a way as general as in Santali is per se noteworthy, the use of proper names as predicates makes a conversion analysis problematic. As a matter of fact, the semantics of all these sentences are uniform, with the differences in meaning arising from the interaction between the semantics of the syntactic positions and the lexical semantics. Hence I am confident in claiming that they can be explained without having to resort to any lexical mechanism.

Obviously, more research is needed regarding the lexical and grammatical properties of proper names in Santali. There are also many open questions about the parts-of-speech system of Santali as the analysis presented in this paper is not suited to argue for a bidirectionality of the flexibility. However, the regular usage of proper names as predicates is a case in point for the generality of the flexibility. If closely related languages with a very similar grammar, such as Mundari, really cannot use proper

names in predicate position, it would be an interesting task for further research to determine where exactly the differences in the lexical and grammatical structure lie.

References

Anderson, J. M. (2007). *The Grammar of Names*. Oxford: Oxford University Press.

Bach, K. (2002). 'Giorgione was so-called because of his name', *Philosophical Perspectives* 16**:** 73–103.

Bhat, D. (1997). 'Noun-verb distinction in Munda languages', in A. Abbi (ed.), *Languages of tribal and indigenous people of India* (MLBD Series in Linguistics 10). Delhi: Motilal Banarsidass, 227–251.

Bodding, P. O. (1926). *Santal Folk Tales. Vol. 1.* Oslo: Aschenhoug & Co.

--- (1927). *Santal Folk Tales*. *Vol. 2.* Oslo: Aschenhoug & Co.

--- (1929a). *Materials for a Santali grammar. Vol. II (mostly morphological)*. Dumka: Santal Mission of the Northern Churches.

--- (1929b). *Santal Folk Tales. Vol. 3.* Oslo: Aschenhoug & Co.

--- (1929-1936). *A Santal Dictionary. Vol. 1-5.* Oslo: Dybwad (on consignment).

Declerck, R. (1988). *Studies on Copular Sentences, Clefts and Pseudo-Clefts* (Series C Linguistica 5). Leuven: Leuven University Press.

Evans, N. and Osada, T. (2005). 'Mundari: The myth of a language without word classes', *Linguistic Typology* 9-3: 351–390.

Frege, G. (1893). 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* 100: 25–50.

Geurts, B. (1997). 'Good news about the description theory of names', *Journal of Semantics* 14: 319–348.

Ghosh, A. (2008). 'Santali', in G. D. S. Anderson (ed.), *The Munda Languages*. London: Routledge, 11-98.

Hengeveld, K. (1992). *Non-verbal predication: theory, typology, diachrony* (Functional Grammar Series 15). Berlin and New York: Mouton de Gruyter.

Hengeveld, K., Rijkhoff, J. and Siewierska, A. (2004). 'Parts-of-speech systems and word order'. *Journal of Linguistics* 40-3: 527–570.

Hengeveld, K. and Rijkhoff, J. (2005). 'Mundari as a flexible language', *Linguistic Typology* 9-3: 406–431.

Higgins, F. R. (1979). *The Pseudo-cleft Construction in English* (Outstanding Dissertations in Linguistics 17). New York: Garland.

Hoffmann, J. (1903). *Mundari Grammar.* Calcutta: Bengal Secretariat Press.

Kneale, W. (1962). 'Modality de dicto and de re', in E. Nagel, P. Suppes and A. Tarski (eds.), *Logic, Methodology and Philosophy of Science. Proceedings of the 1960 International Congress*. Stanford: Stanford University Press, 622–633.

Kripke, S. (1972). 'Naming and Necessity', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*. Dordrecht: Reidel, 253-355, 763-769.

Matushansky, O. (2005). 'Call me Ishmael', in E. Maier, C. Bary and J. Huitink (eds.), *Proceedings of Sub9.* Available at http://www.ru.nl/ncs/sub9

--- (2009) 'On the linguistic complexity of proper names', *Linguistics and Philosophy* 31-5: 573-627.

Mikkelsen, L. (2005). *Copular Clauses: Specification, predication and equation.* (Linguistik Aktuell 85). Amsterdam: Benjamins.

Neukom, L. (2000). 'Argument marking in Santali', *Mon-Khmer Studies* 30: 95–113.

--- (2001). *Santali* (Languages of the World/Materials 323). München: Lincom.
Peterson, J. (2005). 'There's a grain of truth in every "myth", or, why the discussion of lexical classes in Mundari isn't quite over yet', *Linguistic Typology* 9-3, 391–441.
Recanati, F. (1993). *Direct Reference*. Oxford: Blackwell.
Rijkhoff, J. (2003). 'When can a language have nouns and verbs?', *Acta Linguistica Hafniensia* 35: 7–38.
Van Langendonck, W. (2007). *Theory and Typology of Proper Names*. Berlin and New York: Mouton de Gruyter.

Unidirectional flexibility and the noun–verb distinction in Lushootseed*

David Beck

1 Introduction

Recent work on the typology of parts-of-speech systems has shown that a significant parameter of variation in the organization of the lexicon concerns the number of open or major word classes that are recognized in a language. While languages with the familiar Indo-European system distinguish four major 'contentive' classes (noun, verb, adjective, and adverb), it is not uncommon for languages to distinguish fewer. In many such cases, a language with a reduced parts-of-speech inventory conflates two or more major classes, creating a *flexible* part of speech that fills a variety of syntactic roles. One of the most contentious issues that falls out from this observation is whether or not it is possible for a language to conflate all of the major lexical classes, grouping its contentive lexical items into a single, maximally flexible class of words (opposed only by the minor, grammatical classes) and thereby neutralizing the distinction between nouns and verbs. Claims for the absence of a noun–verb distinction have been advanced for a number of languages and are discussed most extensively for languages from the Salishan, Polynesian, and Munda families. Examining these cases reveals that they fall into two general types which I will refer to in this paper, loosely following Evans and Osada (2005), as *precategorial* and *omnipredicative*. The precategorial type of language, as represented by languages of the Munda and Polynesian families, has received the most attention in the recent literature (e.g., Broschart 1991; Croft 2000; Vonen 2000; Hengeveld and Rijkhoff 2005); the omnipredicative type has not been discussed to the same extent, although languages of this kind, particularly those belonging to the Salishan family, are frequently offered uncritically as examples of languages that lack a distinction between nouns and verbs.

In this paper, I will present data from the Salishan language Lushootseed[1] which demonstrates that, while the noun–verb distinction is neutralized in syntactic predicate position, it is still relevant for words used as syntactic arguments, giving us a pattern that will be referred to here as *unidirectional flexibility*. Unidirectional flexibility as the term is used here is intended to complement the notion of 'bidirectional flexibility' put forward by Evans and Osada (2005) as a criterion for determining whether or not a language has genuinely neutralized a part-of-speech distinction. For Evans and Osada, a particular part of speech is considered to be bidirectionally flexible if all of its members can occupy the syntactic roles typical of two (or more) parts of speech, thereby

* The author would like to thank Paulette Levy and Igor Mel'čuk for helping him refine the ideas behind this paper, as well as Jan Rijkhoff, Eva van Lier, and three anonymous reviewers for their helpful critiques and suggestions. My thanks also to the late Thom Hess for providing me with the data and understanding of Lushootseed that lie behind this paper. None of the above bear any responsibility for my errors.

[1] Lushootseed is a member of the Central Coast Salish branch of the Salishan family of languages, and was formerly spoken throughout the Puget Sound region of Washington State. It is currently the native language of no more than a handful of elders. Lushootseed data not cited as being from published sources are drawn from my own textual database built from materials collected by Thomas M. Hess; these sources are cited by speaker's initials, title of text, and line number. Re-glossing and reanalysis of some examples from older published sources has occurred in a few cases for the sake of consistency, and to bring them into line with the conventions followed in Beck (in progress) and Beck and Hess (n.d.).

conforming to the definitions of both lexical classes. Unidirectional flexibility, on the other hand, entails that for a particular pair of lexical classes, X and Y, X can appear in the syntactic roles criterial for Y, but Y cannot appear in the roles criterial for X. While unidirectional flexibility entails the neutralization of a parts-of-speech distinction in a particular syntactic environment, it cannot be equated with the complete absence of the distinction.

Unidirectional flexibility and methods for establishing it will be discussed in Section 2 of this paper, following which Lushootseed data illustrating this pattern will be presented in Section 3. The facts in Lushootseed closely parallel those of other Salishan languages, which in turn seem substantially the same as the patterns seen in other languages that have been claimed to follow the omnipredicative pattern of noun–verb neutralization, implying that omnipredicative languages in general show only unidirectional, rather than genuine bidirectional, flexibility between nouns and verbs. Since precategorial languages are also argued (for different reasons) by Evans and Osada (2005) not to represent a genuine example of noun–verb flexibility, it seems probable that a distinction between nouns and verbs is indeed a universal of human language (Croft 2003). Evans and Osada's position, some counter-arguments to it, and some of the more general implications of this discussion for typological approaches to parts-of-speech systems will be discussed in Section 4.

2 Flexibility in parts-of-speech systems

The notion of flexibility in parts-of-speech systems is first articulated in the context of a full typology of lexical classes by Hengeveld (1992a, b), who uses the term 'flexible' to refer to a part of speech that meets two or more of the definitions for lexical classes given in (1), these definitions hinging crucially on the morphosyntactic properties of classes of lexical items appearing in certain criterial syntactic environments:[2]

(1) *verb*: a lexical item which, without further measures being taken, has predicative use only
noun: a lexical item which, without further measures being taken, can be used as a syntactic argument
adjective: a lexical item which, without further measures being taken, can be used as the modifier of a noun
adverb: a lexical item which, without further measures being taken, can be used as the modifier of a syntactic predicate[3]
(adapted from Hengeveld 1992b:58)

Thus, a class of lexical items is said to be flexible if it meets, say, both the definition of an adjective and of an adverb simultaneously. Hengeveld further proposes, based on a

[2] Note that I have re-formulated the definitions, which are couched in the terminology of Functional Grammar (Dik 1997), using more neutral descriptive terms for syntactic environments. I will continue to follow this practice throughout the remainder of this discussion.

[3] Adverbs in many languages can modify other elements such as adjectives and other adverbs as well, but this is by no means universal. Jesperson (1965) also notes that adverbs in English do not modify nominal syntactic predicates; once again, this is not universal, but the implications of this for this definition of adverbs merit some consideration.

moderately large sample of languages, that the patterns of flexibility thus defined are not unconstrained, but follow the implicational hierarchy shown in (2):

(2) Parts of Speech Hierarchy (adapted from Hengeveld, Rijkhoff and Siewierska 2004)
Syntactic predicate > Syntactic argument > Adnominal modifier > Adverbal modifier

According to (2), a part-of-speech system that has a class of words used both as unmarked syntactic predicates and as unmarked syntactic arguments will also use the same class of words for adnominal and adverbal modification; a flexible class of words that is used as a syntactic argument and adnominal modifier must also be flexible with respect to adverbal modification; and so on. The resulting taxonomy of flexible parts-of-speech systems is shown in Fig. 1:

		SYNTACTIC ROLE			
Part-of-speech systems		SYNTACTIC PREDICATE	SYNTACTIC ARGUMENT	ADNOMINAL MODIFIER	ADVERBAL MODIFIER
Flexible systems	Type 1	Contentive			
	Type 2	Verb	Non-verb		
	Type 3	Verb	Noun	Modifier	

Fig. 1: Taxonomy of flexible parts-of-speech systems (adapted from Hengeveld 1992a)

Since its inception, this taxonomy has been influential and controversial, both in terms of the typological predictions it makes and in terms of the methodological implications it has for the investigation of lexical class systems.

One methodological question that is of importance to us in the context of the present discussion is the notion of 'without further measures', which is defined by Hengeveld in rather vague terms and seems to correspond roughly to additional morphological or syntactic means required for the use of a particular lexical item in a non-canonical syntactic role (referred to by Tesnière 1934, 1959 as 'transfer'). Some of the implications of this are discussed in Beck (2002), where it is proposed that 'further measures' be re-defined in terms of the relative markedness of particular lexical classes of item in specific syntactic roles. Of the criteria for determining relative markedness, the most relevant for this paper is the notion of Structural Complexity:

(3) **Structural Complexity**: An element X is marked with respect to another element Y if X is more complex, morphologically or syntactically, than Y

Applying this measure to parts-of-speech typology, establishing the markedness of a lexical class X relative to lexical class Y requires showing that members of Class X are relatively more structurally complex than those of Class Y in a criterial syntactic environment A. This is in essence the equivalent of the descriptive claim that words of Class X are the target of morphosyntactic rules (aimed specifically at Class X, which must therefore be recognized in the lexicon) allowing for their use in environment A. Re-formulating this in terms of markedness allows the analyst a principled way to establish language-specific diagnostics and criteria for structural complexity (thereby avoiding what Croft 2005: 434 refers to as 'methodological opportunism'). When a part-of-speech distinction exists between two word classes, each with its own distinct

unmarked syntactic role, the comparison of the properties of Classes X and Y in two of the criterial syntactic roles identified in (1), A and B, would give us the pattern shown in Fig. 2:

	ROLE A	ROLE B
CLASS X	*marked*	*unmarked*
CLASS Y	*unmarked*	*marked*

Fig. 2: Bidirectional lexical class distinction

Here, Class X (say, for English, nouns) is relatively more complex in terms of derivational or syntactic means, and therefore marked, in Role A (syntactic predicate) than Class Y (verbs), while in Role B (syntactic argument) Class Y is relatively more complex than Class X, giving us a clear bidirectional lexical class distinction.[4]

In the case of a truly flexible part of speech, we would expect that for an established class of words, any bipartition of that class into two putative sub-classes X and Y (at random or based on semantic criteria) would show the pattern in Fig. 3:

	ROLE A	ROLE B
CLASS X	*unmarked*	*unmarked*
CLASS Y	*unmarked*	*unmarked*

Fig. 3: Bidirectional flexibility

In this case, no criteria for relative structural markedness can be found that distinguish between Class X and Class Y in criterial Role A or between Class X and Class Y in criterial Role B (that is, there is no lexical class distinction between the two sets). This situation corresponds to what Evans and Osada (2005) refer to as 'bidirectionality', and contrasts with the situation illustrated in Fig. 4, which might be termed 'unidirectional' flexibility:

	ROLE A	ROLE B
CLASS X	*unmarked*	*unmarked*
CLASS Y	*unmarked*	***marked***

Fig. 4: Unidirectional flexibility

In this situation, words belonging to Class X are unmarked in both criterial Roles A and B, whereas words in Class Y are unmarked only in Role A, but are marked in Role B — in effect, the distinction between Classes X and Y is present in the language but is 'neutralized' for Role A. Situations such as this are not uncommon in languages, but do not constitute genuine flexibility: it is still possible to define distinct word-classes based on the contrastive properties of the two classes in Role B. Thus, for instance, in a

[4] The same type of argumentation can, of course, be made on the basis of markedness established by other criteria.

language where words with substantive meanings (Class X) are both unmarked as predicates (Role A) and as syntactic arguments (Role B), but words expressing events (Class Y) are unmarked predicates but marked arguments, Class Y still conforms to the definition of 'verb' given in (1) but does not conform to the definition of 'noun'. Class X, on the other hand, conforms to both (or would, without the stipulation that verbs be 'only' syntactic predicates), but, given the contrast with Class Y, can be classified as a flexible class of nouns. As will be shown in Section 2 below, the apparent neutralization of the noun–verb distinction in Salishan languages constitutes a clear case of this type of unidirectional flexibility; because Salishan provides a typical case of what Evans and Osada (2005) refer to as an 'omnipredicative' language (borrowing the term from Launey 1994) the discussion below strongly suggests that languages in this category do not constitute a genuine case of noun–verb flexibility.

Another controversial aspect of the definitions of parts of speech in (1) is the absence of semantic criteria associated with any of the word classes (Beck 2002). This seems unfortunate from a theoretical point of view, given the well-known and quite robust clusterings of certain meaning-types around certain parts of speech, shown in Fig. 5 (cf. Croft 1991):

Substantives (people, place, thing)	Events (action, process, state)	Property concepts (dimension, age, value, etc.)
Noun	Verb	Adjective

Fig. 5: Proto-typical associations of meaning-type and lexical classes

While it is well-known that semantic category membership is (at best) problematic for establishing lexical-class membership, it is nevertheless true that accounting for these patterns is a desirable feature for a parts-of-speech typology. One of the unintended consequences of this focus on syntactic over semantic criteria is that it often leads to a tacit methodological bias towards strictly morphosyntactic comparisons of related wordforms in different syntactic environments without attention to concomitant differences in their meanings (a similar point with respect to the Salishan noun/verb issues is made in Van Eijk and Hess 1986: 328). In cases where the two wordforms being compared are phonologically identical, however, inattention to semantics — particularly changes in the meaning of one of the forms associated with its appearance in a particular syntactic role — leaves the door open to a (mis)analysis wherein two words are judged to be morphosyntactically equivalent in spite of a significant semantic difference between them. This approach begs the question of whether or not the two items being compared are, in fact, the same word with a flexible distribution, rather than two homophonous forms, each with its own meaning and syntactic distribution. While such cases generally pass without comment in languages like English, which has a number of such pairs of homophonous forms (e.g., *hammer*$_{N}$ vs. *hammer*$_{V}$, *cook*$_{N}$ vs. *cook*$_{V}$), there appear to be languages (the most frequently cited examples being languages from the Munda and Polynesian families) where such pairs are very common, perhaps to the point of constituting the bulk of the lexicon. As Evans and Osada (2005) note, languages of this kind, often referred to as 'precategorial' languages, constitute a second language-type that is often analyzed as having a flexible class of words fitting

the definition of both nouns and verbs.[5] For the case of Mundari, Evans and Osada argue (on different grounds) that this situation is, like the omnipredicative case, not an example of true flexibility. Thus, with both putative types of noun–verb flexibility in doubt, it would seem that the case for the typology in Fig. 1 is considerably weakened, at least insofar as the possibility of having a language with a single major part of speech is concerned. I will return to this point in the conclusion to this paper.

3 Unidirectional flexibility: Noun and verb in Lushootseed

One of the most frequently-cited cases of a language family that is flexible with respect to the noun–verb distinction is that of Salishan languages. These claims were put forth initially in the specialist literature (e.g., Kuipers 1968; Kinkade 1983; Jelinek and Demers 1994) and then adopted by typologists interested in variation in parts of speech systems (e.g., Broschart 1991; Sasse 1993; Bhat 1994; Hengeveld and Rijkhoff 2005), although the current consensus in the Salishanist community seems to be against this position (e.g., van Eijk and Hess 1986; Demirdache and Matthewson 1995; Matthewson and Davis 1995; Davis and Matthewson 1998, 1999; Kroeber 1999; Beck 2002). The primary argument for the absence of a noun–verb distinction in Salishan is data such as that from Lushootseed shown in (4):[6]

(4)a. *sbiaw ti ʔuǩʷ*

sbiaw ti ʔuǩʷ
coyote SPEC go
'the one who goes is Coyote'

(Van Eijk and Hess 1986: 324)

[5] In actual fact, Evans and Osada list four types of putative noun–verb flexibility. In addition to the omnipredicative and precategorial type, they mention the 'Broschartian' language and languages with 'rampant' conversion. The thrust of their article, however, is to show (I believe correctly) that the precategorial and Broschartian types of language are, in fact, better analyzed as languages with rampant conversion — and that this last category does not in fact represent a true example of noun–verb flexibility. Since the term 'rampant conversion' language presupposes the outcome of this discussion, I have chosen to use 'precategorial' for the moment as a more neutral cover term.

[6] The abbreviations used in this paper are as follows: – = morpheme boundary; = = clitic boundary; • = lexical suffix boundary; 1, 2, 3 = first-, second-, third-person; ADD = additive; ADNM = adjunct nominalizer; ALTV = allative; ATTN = attenuative; CLS = classifier; CNTRFG = centrifugal; CNTRPT = centripetal; CTD = contained; DAT = dative; DC = diminished control; DEF = definite; DET = determiner; DIST = distal; DMA = demonstrative adverb; DSD = desiderative; DSTR = distributive; ECS = external causative; FEM = feminine; FOC = focalizer; HAB = habitual; HMN = human; ICS = internal causative; INCH = inchoative; INT = interrogative; INTJ = interjection; IRR = irrealis; NEG = negative; NM = nominalizer; NSPEC = non-specific; OBJ = object; PASS = passive; PFV = perfective; PL = plural; PO = possessive; PR = preposition; PROG = progressive; PROP = propriative; PROX = proximal; PRTV = partitive; PTCL = particle; REM = remote; SBJ = subjunctive; SG = singular; SCONJ = sentential conjunction; SPEC = specific; STAT = stative; SUB = subject, UNQ = unique.

b. *p'q'adᶻəxʷ tiʔəʔ ʔəxʷx̌qabac*

p'q'adᶻ = əxʷ tiʔəʔ ʔ əs–dxʷ–x̌q•abac
rotten.log=now PROX STAT–CTD–wrapped•body
'what is wrapped up in it is a rotten log'

[HM Star Child, line 52]

c. *t'əq'ʷ tiʔił x̌ʷaqʷabac*

t'əq'ʷ tiʔił x̌ʷaqʷ•abac
snap DIST wrapped•body
'what was wrapped around her waist snapped'

[DS Star Child, line 134]

d. *ʔəbil'čəxʷ łušudxʷ tiʔił łuləgʷax̌ʷ …*

ʔəbil' čəxʷ łu = šuɬ–dxʷ tiʔił łu = lə = gʷax̌ʷ
if 2SG.SUB IRR=see–DC DIST IRR=PROG=walk
'if you see someone travelling …'

(Hess 2006: 49, 180)

Like other Salishan languages, Lushootseed is strongly predicate-initial. Sentences such as (4a) represent a fairly common type of construction where the syntactic predicate is a word with a substantive meaning, *sbiaw* 'Coyote', whose subject appears to be a word expressing an event, *ʔux̌ʷ* 'go'. As in all sentences with substantive syntactic predicates, the meaning of the construction here is equative. Likewise, in (4b) the syntactic predicate is the substantive *p'q'adᶻ* 'rotten log' and has as a subject what appears to be the translation equivalent of a verb, *dxʷx̌qabac* 'be wrapped up inside', inflected for the stative aspect and preceded by a determiner. As shown in (4c), words corresponding to English verbs are not confined to argument-position in constructions with substantive predicates: in this sentence, the predicate is *t'əq'ʷ* 'snap', but the subject is apparently the expression of an event *x̌ʷaqʷabac* 'be wrapped around body'. Likewise, (4d) shows that event-words can be direct objects of transitive predicates. Such words can also appear in other syntactic argument roles such as agentive complement of a passive (5a), as well as being found in (non-criterial) roles such as complement of a preposition (5b) which are cross-linguistically most typical of nouns:

(5)a. *diʔł kʷi sgʷəgʷaʔtubs* ***ʔə kʷədiʔ ʔugʷəgʷaʔtxʷ***

diʔł kʷi s = gʷə–gʷaʔ–txʷ–b = s ʔə kʷədiʔ
sudden REM NM = ATTN–accompanied–ECS–PASS = 3PO PR that.one
ʔu–gʷə–gʷaʔ–txʷ
PFV–ATTN–accompanied–ECS
'suddenly she was joined by the one who accompanied her'
(lit. 'her being joined by that one who accompanied her was sudden')

[DS Star Child, line 76]

b. *gʷəl ləxʷəbtəb əlgʷəʔ* ***dxʷʔal ʔəsqʼil***

gʷəl	lə=xʷəb–t–b	əlgʷəʔ	dxʷ–ʔal	ʔəs–qʼil
SCONJ	PROG=thrown–ICS–PASS	PL	CNTRPT–at	PFV–aboard

‘and they were thrown aboard’

(Hess 2006: 58, line 399)

In fact, if judged on superficial distributional criteria, there appear to be no syntactic roles that are open to words with substantive meanings (i.e., words we would expect to be nouns) that are not also open to words that express events (words we would expect to be verbs).

On the basis of evidence such as that presented in (4) and (5), it would seem that a *prima facie* case for the absence of a noun and verb distinction in Salishan languages can be made. However, closer examination of data from Lushootseed reveals that all in fact is not as it seems: while it may be true that the noun–verb distinction is neutralized in syntactic predicate position (Section 3.1), it can be shown to persist for words from these two classes in syntactic argument position (3.2). The primary piece of evidence for this persistence is that argument-phrases like *ti ʔuǰʷ* ‘the one who goes’ in (4a) are, as their translation implies, headless relative clauses (3.2.1). Furthermore, there are constructions in which words expressing events appearing in argument position show clear morphological and syntactic evidence of recategorization as nominals (3.2.2), as well as constructions in which the treatment of particular words depends crucially on whether they belong to a nominal or a verbal lexical class (3.2.3). Thus, Lushootseed (and most likely Salishan languages in general) do not meet Evans and Osada’s (2005) criteria of bidirectionality, and so do not constitute a true case of noun–verb flexibility, but instead correspond to a case of unidirectional flexibility.

3.1 Neutralization in predicate position

As shown in the previous section, word classes in Lushootseed do show flexibility in that the noun–verb distinction in the lexicon is neutralized in syntactic predicate position, as illustrated by the sentences in (4a) and (b), which have as syntactic predicates words with substantive meanings (‘coyote’ and ‘rotten log’, respectively). As noted earlier, Lushootseed is a predominantly predicate-initial language, and so in the matrix clause the syntactic predicate is the first word belonging to a major word class:

(6)a. *ʔuǰʷəxʷ tiʔəʔ sgʷəlub*

ʔuǰʷ=əxʷ	tiʔəʔ	sgʷəlub
go=now	PROX	pheasant

‘Pheasant goes now’

(Hess 1998: 79, line 40)

b. *kʷədatəb tiʔiɬ*

kʷəda–t–b	tiʔiɬ
taken–ICS–PASS	DIST

‘that one was taken’

(Hess 2006: 59, line 428)

c. *sx̌aʔhus tsiʔəʔ čəgʷas diič'uʔ*

sx̌aʔhus tsiʔəʔ čəgʷas diič'uʔ
sawbill PROX:FEM wife one:HMN
'one of the wives is Sawbill'

(Hess 2006: 22, line 5)

d. *sʔuladxʷ tiʔiɬ*

sʔuladxʷ tiʔiɬ
salmon DIST
'that is a salmon'

(Hess and Hilbert 1976: vol. I, 7)

As shown here, overt subjects immediately follow the predicate phrase both in sentences predicated on words expressing events and in sentences with substantive predicates. Lexical subjects may be full argument phrases as in (6a) and (6c), or simply demonstrative determiners as in (6b) and (d). Subject inflection in matrix clauses without NP subjects is marked by subject markers, the full paradigm for which is given with *ʔux̌ʷ* 'go' in (7):

(7)a. *ʔux̌ʷ čəd*

ʔux̌ʷ čəd
go 1SG.SUB
'I go' 'we go'

b. *ʔux̌ʷ čəɬ*

ʔux̌ʷ čəɬ
go 1PL.SUB

c. *ʔux̌ʷ čəxʷ*

ʔux̌ʷ čəxʷ
go 2SG.SUB
'you$_{SG}$ go'

d. *ʔux̌ʷ čələp*

ʔux̌ʷ čələp
go 2PL.SUB
'you guys go'

e. *ʔux̌ʷ*

ʔux̌ʷ Ø
go 3SUB
'he/she/it/they go'

These same subject markers are also used with substantive predicates, as in (8):

(8) *ʔaciɬtalbixʷ čəd*

ʔaciɬtalbixʷčəd
Indian 1SG.SUB
'I am an Indian'

(Hess and Hilbert 1976: vol. I, 36)

Subject markers immediately follow the predicate, preceding any objects (9a), but only as long as the predicate is clause-initial; otherwise, the subject marker migrates to

sentence-second position, immediately following any predicate modifiers as in (9b) and (9c):

(9) a. *ług^{w}iid čəd k^{w}si sx^{w}iʔuq'^{w}*

łu = g^{w}ii–d čəd k^{w}si sx^{w}iʔuq'^{w}
IRR=call–ICS 1SG.SUB REM:FEM Basket.Ogress
'I'll call the Basket Ogress'

[AJ Basket Ogress, line 27]

b. *cick'^{w} čəd ʔəx^{w}ʔuǩ^{w}əb*

cick'^{w} čəd ʔəs–dx^{w}–ʔuǩ^{w}–ab
very 1SG.SUB STAT–CTD–go–DSD
'I very much want to go'

(Hess 1995: 90)

c. *łuhik^{w} čəd stubš łuluƛ'iləd*

łu = hik^{w} čəd stubš łu = luƛ'–il = əd
IRR=big 1SG.SUB man IRR=old–INCH=1SG.SBJ
'I will be a big man when I grow old'

(Bates, Hess and Hilbert 1994: 109)

As the data in (9) show, migration applies equally to the subjects of event-word and substantive predicates.

The examples in (9a) and (9c) illustrate another common source of confusion about the distinction between lexical classes is the existence of phrase-level inflectional or quasi-inflectional proclitics marking tense and mood. This set of proclitics includes the irrealis *łu=* seen in (9a) and (9c), the subjunctive clitic *g^{w}ə=* (11a), the additive *bə=* (11b), and the habitual *ƛ'u=* (30b). These proclitics tend to appear on the first 'contentive' word (as opposed to clitic, particle, determiner, etc.) in a phrase, whatever its lexical class, as illustrated by the fifth member of this set, the past tense marker *tu=*:

(10) a. *tuk^{w}aʔtəbəx^{w}*

tu = k^{w}aʔ–t–b = əx^{w}
PAST=released–ICS–PASS=now
'he was let go'

(Hess 2006: 20, line 220)

b. *tuq'iyaƛ'əd tiʔiɬ tusč'istx^{w}s*

tu = q'iyaƛ'əd tiʔiɬ tu = sč'istx^{w}–s
PAST=slug DIST PAST=husband–3PO
'Slug had been her husband'

(Hess 1998: 70, line 135)

c. *huy čəxʷ tascutəb ʔə tiʔiɬ tabsbəda ʔ*

huy	čəxʷ	tu = ʔas–cut–t–b	ʔə	tiʔiɬ
SCONJ	2SG.SUB	PAST = STAT–say–ICS–PASS	Pr	DIST

tu = ʔas–bəs–bəda ʔ

PAST = STAT–PROP–offspring

'for you were told to by the (deceased) one who had a daughter'

(Hess 1998: 98, line 203)

d. *diɬ tuxʷul'tubəɬa ʔ*

diɬ	tu = xʷul'	tu = bə = ɬa ʔ
FOC	PAST=just	PAST=ADD=arrive

'it was he who had just kept arriving there'

(Hess 2006: 21, line 244)

These clitics can appear on event-word predicates, as shown in (10a), or on substantive arguments, as in (10b) and (c). (10b) also shows the past-tense clitic appearing on a substantive predicate, *q'iyaχ'əd* 'slug'.[7] The phrase-level clitics may appear more than once in a clause, (10b) and (c), or even be iterated within a single phrase, (10d). The fact that all of these clitics are found attached to all types of words in both predicate and argument phrases has been in some measure responsible for the misapprehension that lexemes of all classes take word-level inflections for tense, aspect, and mood categories.[8]

Just as subject inflection is the same for substantive and event-word predicates in matrix clauses, it is also the same in subordinate subjunctive clauses whose syntactic predicates are words of either type. These clauses take a special series of subject enclitics, illustrated in (11) by *=əs* '3 subjunctive':

(11) a. *ʔəsxəc gʷəxʷit'iləs əlgʷəʔ*

ʔəs–xəc	gʷə = xʷit'–il = əs	əlgʷəʔ
STAT–afraid	SBJ=descend–INCH=3SBJ	PL

'he is afraid they will fall'

(Hess 1967: 76)

[7] Further discussion of the issue of tense in Salishan languages can be found in Bates (2002), Burton (1997), Matthewson (2002, 2005), and Wiltschko (2003).

[8] There is in fact an inflectional difference between substantives and words expressing events — namely, that the latter take aspectual inflections, such as the stative aspect-marker *as-* in (10c), whereas the former do not. However, this distinction is true in all environments, not just in predicate position, and so the absence of aspectual inflections on substantive predicates cannot be used as a measure of their (i.e. substantives') markedness in this syntactic environment.

b. *xʷɬub bəhikʷ tiʔiɬ bəshuyitəbs əlgʷəʔ stuləkʷ, gʷəstuləkʷəs*

xʷɬub	bə = hikʷ	tiʔiɬ	bə = s = huy–yi–t–b = s	əlgʷəʔ
ultimately	ADD = big	DIST	ADD = NM = made–DAT–ICS–PASS = 3PO	PL
stuləkʷ	gʷə = stuləkʷ = əs			
river	SBJ = river = 3SBJ			

'finally an even bigger river was made for them, if it was a river'

(Hess 2006: 36, line 354)

The full set of subjunctive person-markers is given in Fig. 6:

	SG	PL
1	*=ad/=əd*	*=aɬi/=əɬi*
2	*=axʷ/=əxʷ*	*=aləp/=ələp*
3	*=as/=əs*	

Fig. 6: Subjunctive person-markers

These subjunctive person-markers are enclitics that become phonologically dependent on the immediately preceding word, irrespective of its lexical class:

(12) *gʷəckʼʷaqidələp gʷučaləc*

gʷə = ckʼʷaqid = ələp	gʷə = ʔu–čala–t–s		
SBJ=always=2PL.SBJ	SBJ=PFV–chased–ICS–1SG.OBJ		

'if you folks always chase me'

(Hess 1967: 52)

Here, the subjunctive person marker appears in a clause introduced by an adverb, *ckʼʷaqid* 'always'; because the adverb is clause-initial, the person-marker is encliticized to this word rather than to the syntactic predicate, maintaining its clause-second position. Once again, the fact that these person inflections are sentence-second clitics is often overlooked, and these morphemes have been used as evidence for the claim that words of any class (including adverbs) can be inflected for person and number of their subjects in subjunctive clauses.

As noted by Kinkade (1983) for Salishan in general, the patterns shown by words expressing events and substantives in predicate position also apply to other types of word. Thus, Lushootseed has clauses predicated on a variety of word classes such as lexical pronouns (13), adverbs (14), numerals (15), and interrogative words (16):

(13) *ʔəca kʷi ɬuɬiɬičʼid tiʔiɬ tatačulbixʷ*

ʔəca	kʷi	ɬu = ɬi–ɬičʼi–d	tiʔiɬ	tatačulbixʷ	
I	REM	IRR=ATTN–cut–ICS	DIST	big.game	

'the one who will cut up the big game animal is me'

[DS Star Child, line 304]

(14) *tudiʔ tə dukʷibəɬ*

tudiʔ tə dukʷibəɬ
over.there NSPEC Changer
'Changer is way over there'

(Hess 1995: 81, ex. 6)

(15) *saliʔ kʷi ɬuʔəƛ̓'txʷ čəxʷ č'ƛ̓'aʔ*

saliʔ kʷi ɬu=ʔəƛ̓'–txʷ čəxʷ č'ƛ̓'aʔ
two REM IRR=come–ECS 2SG.SUB stone
'you will bring two stones' (lit. 'the stones that you will bring are two')

[AW Basket Ogress, line 80]

(16) a. *tučadəxʷ čəxʷ*

tu=čad=əxʷ čəxʷ
PAST=where=now 2SG.SUB
'where have you been?'

b. *tul'čad čəxʷ*

tul'–čad čəxʷ
CNTRFG–where 2SG.SUB
'where are you coming from?'/'where are you from?'

(Bates et al. 1994: 59)

Kroeber (1999) also notes that many Salishan languages, including Lushootseed, allow prepositional phrases to head clauses, as in (17):

(17) a. *dxʷʔal tə hud tə sxʷit'il ʔə tə biac*

dxʷ–ʔal tə hud tə s=xʷit'–il ʔə tə biac
CNTRPT–at NSPEC fire NSPEC NM=descend–INCH PR NSPEC meat
'into the fire falls the meat' (lit. 'the fall of the meat is into the fire')

(Kroeber 1999: 381)

b. *tul'ʔal čəd sqaǰət*

tul'–ʔal čəd sqaǰət
CNTRFG–at 1SG.SUB Skagit
'I am from Skagit'

(Bates et al. 1994: 6)

Note that in (17b) the subject marker immediately follows the preposition, separating it from its complement in order to maintain sentence-second position.

Predicates such as those in (13)–(17) also take the subjunctive person-markers in subjunctive subordinate clauses:

(18) a. *gʷuda?atəb dᶻəɬ gʷəcədiɬəs kʷi gʷəu?atəbəd*

gʷə=?u–da?a–t–b dᶻəɬ gʷə=cədiɬ=əs kʷi gʷə=?u–?atəbəd
SBJ=PFV–named–ICS–PASS PTCL SBJ=him/her=3SBJ REM SBJ=PFV–die
'it seems that the one who died should be named!'

(Hess 1998: 71, line 158)

b. *gʷəl wiliq'ʷitəb tutul'čadəs*

gʷəl wiliq'ʷi–t–b tu=tul'–čad=əs
SCONJ ask–ICS–PASS PAST=CNTRFG–where=3SBJ
'and they asked him where he might be from'

(Hess 1998: 97, line 166)

c. *… ?aləs tadi? siq'gʷas ?ə tə šəgʷɬ*

?al=əs tadi? siq'gʷas ?ə tə šəgʷɬ
at=3SBJ over.there bifurcation PRN SPEC path
'… where there is a fork in the path over there'

[AW Basket Ogress, line 99]

(18a) shows the third-person subjunctive clitic attached to a lexical pronoun, (18b) shows it cliticized to an interrogative word, and in (18c) it is bound to a preposition, meaning that the putative neutralization of the noun–verb distinction in predicate position applies to a much broader range of lexical items than simply to nouns and verbs – hence, the claim in Kinkade (1983) that the Salishan lexicon distinguishes only 'predicate' and particles, the former class subsuming nouns, verbs, and anything else that is an eligible syntactic predicate.

Based on the data presented here, then, it does seem that Lushootseed (like most of the languages in the family) shows a great deal of flexibility with respect to what class or classes of word are unmarked syntactic predicates. Nevertheless, the neutralization of lexical class distinctions in predicate position is not enough to demonstrate that the noun–verb distinction is non-existent in the Lushootseed lexicon or is irrelevant to Lushootseed syntax: in addition to showing that all words are unmarked in the criterial syntactic role for verbs, it is also necessary to show that all words are unmarked in the criterial syntactic role for nouns. As will be seen in the following section, this is clearly not the case.

3.2. Contrastive behaviour in argument position

As demonstrated in the preceding section, the syntactic parallels between substantive and event-word predicates in Lushootseed are exact, and the similarities between the two are heightened by the behaviour of person-markers and phrase-level inflectional clitics, which contribute to the illusion that the inflections pertaining to the predicate phrase are morphological categories of a conflated noun–verb lexical class.[9] This pattern of neutralization in predicate position, however, is not all that uncommon, and is

[9] See also Jacobsen (1979), who makes the same observation about the putative noun–verb neutralization in Nootka, another omnipredicative language.

found in a variety of languages (e.g., Buriat, Arabic, Nanay, Beja) where the noun–verb distinction is not in question (Beck 2002: 107–108). Although bare nominal predicates are allowed in such languages, the distinction between nouns and verbs is rarely in doubt once other types of evidence are taken into consideration. In Lushootseed, the crucial evidence is found by examining the syntax of argument phrases: of the open lexical classes, only words with substantive meanings are unmarked syntactic arguments. Consider the following:

(19) a. *tiləb ʔuʔəƛ̓ tiʔəʔ qaw'qs*

tiləb ʔu–ʔəƛ̓ tiʔəʔ qaw'qs
immediately PFV–come PROX raven
'right away Raven showed up'

[MW Star Child, line 101]

b. *gʷəl dił łučəgʷas tiʔił łučəbaʔtxʷ tiʔił*

gʷəl dił łu=čəgʷas tiʔił łu=čəbaʔ–txʷ tiʔił
SCONJ FOC IRR=wife DIST IRR=pack–ECS DIST
'and so the one who (can) carry that will be (my) wife'

[MW Star Child, line 77]

c. *x̌ʷul' buusał kʷi sp'ic'ids*

x̌ʷul' buus•ał kʷi s=p'ic'i–d=s
just four•CLS REM NM=wrung.out–ICS=3PO
'just four times she wrings it out' (lit. 'her wringing it out is just four times')

[HM Star Child, line 66]

On the surface, the subjects of the sentences in (19) all appear to have a similar syntactic structure: a determiner followed by a lexical item with either a substantive meaning (19a) or a meaning expressing an event (19b–c). However, closer examination of the syntactic properties of argument-phrases based on event-words reveals that constructions of the type shown in (19b) are, in fact, headless relative clauses (Section 3.2.1), while those in (19c) are non-finite clauses resembling English gerunds (3.2.2): in other words, when event-words are used in argument position, they are marked in terms of Structural Complexity, requiring us to make a distinction between these words (marked syntactic arguments — i.e., verbs) and substantives (unmarked syntactic arguments — i.e., nouns). Further evidence for the necessity of maintaining this distinction can be found in other non-criterial syntactic environments such as negative constructions (3.2.3), where the interpretation and syntactic treatment of complements of the negative predicate depends on whether the complement is a word with a substantive meaning (a noun) or a word expressing an event (a verb). This evidence strongly supports the conclusion that any flexibility in the Lushootseed lexicon that neutralizes the distinction between noun and verb applies only to the syntactic role of predicate, but is unidirectional and does not apply to the role of syntactic argument.

3.2.1 Headless relative clauses

In spite of the frequent appearance of event-words in argument position in sentences such as (19b) above, these constructions can be shown to be marked in the sense of being syntactically more complex than ordinary nominal arguments: they are, in fact, headless relative clauses (Beck 2002: 113–122).[10] The best evidence for this comes from the restrictions on accessibility to relativization (in the sense of Keenan and Comrie 1977) that hold for both nominally-headed and headless relative clauses: all else being equal, in clauses with a third-person subject and a third-person object, only the subject can be relativized:

(20) a. *ʔušudxʷ čəɬ* ***ti č'ač'as ʔutəsəd tiʔiɬ stubš***

ʔu–šuɬ–dxʷ čəɬ ti č'ač'as ʔu–təs–əd tiʔiɬ stubš
PFV–see–DC 1PL.SUB SPEC child PFV–hit–ICS DIST man
'we saw the boy that hit the man'
*'we saw the boy that the man hit'

b. *ʔušudxʷ čəd* ***ti sqʷəbayʔ ʔuč'axʷatəb ʔə tiʔiɬ č'ač'as***

ʔu–šuɬ–dxʷ čəd ti sqʷəbayʔ ʔu–č'axʷa–t–b ʔə tiʔiɬ
PFV–see–DC 1SG.SUB SPEC dog PFV–clubbed–ICS–PASS PR DIST
č'ač'as
child
'I see the dog that was clubbed by the boy'

(Hess and Hilbert 1976: II, 124–125)

(20a) gives an example of a modifying relative clause with a third-person subject and a third-person object; in this case, the only interpretation of the sentence possible is that of a subject- relative clause. When the object- reading is desired, it is necessary to passivize the embedded clause as in (20b). The same holds for headless relative clauses such as those shown in (21):

(21) a. *wiw'su* ***tiʔəʔ ʔučalad tiʔəʔ sqʷəbayʔ***

wiw'su tiʔəʔ ʔu–čala–d tiʔəʔ sqʷəbayʔ
children PROX PFV–chased–ICS PROX dog
'the ones who chased the dog are the children'
*'the ones who the dog chased are the children'

b. *sqʷəbayʔ* ***ti ʔučalatəb ʔə tiʔiɬ wiw'su***

sqʷəbayʔ ti ʔu–čala–t–əb ʔə tiʔiɬ wiw'su
dog SPEC PFV–chased–ICS–PASS PR DIST children
'the one who is chased by the children is the dog'

(Hess 1995: 99)

[10] The term 'headless' here should be taken to refer only to the absence of a substantive (nominal) head modified by the clause—the constructions are, in fact, headed syntactically by the preceding determiners. See Kroeber (1999: 258–261) for a general discussion of the construction in the family.

In (21a), the only interpretation open to the sentence is the one where the headless relative clause identifies the subject of the embedded verb, in spite of the fact that the opposite interpretation, where the dog chases the children, is semantically and pragmatically quite plausible. Again, where this is the desired interpretation, the embedded clause appears in the passive, as in (21b).[11]
When the subject of the embedded clause is first- or second-person, only object- relative clauses are possible:

(22) a. *ʔušudxʷ čəɬ* ***ti č'ač'as ʔutəsəd čəd***

ʔu–šuɬ–dxʷ čəɬ ti č'ač'as ʔu–təs–d čəd
PFV–see–DC 1PL.SUB SPEC child PFV–hit–ICS 1SG.SUB
'we saw the boy that I hit'

(Hess and Hilbert 1976: II, 125)

b. *skəyu təɬ* ***tiʔiɬ ʔucucuuc čələp***

skəyu təɬ tiʔiɬ ʔu–cut–cut–c čələp
ghost truly DIST PFV–DSTR–say–ALTV 2PL.SUB
'what you guys are talking about is truly a ghost'

(Hess 1998: 94, line 107)

This is almost certainly a syntactic restriction, as the subject-markers are not themselves nominals and so cannot head an NP or be modified by a relative clause. There are no examples of first- or second-person pronouns heading a relative clause construction, but there are numerous examples of pronouns functioning as predicates of sentences with headless relative clause subjects. In these cases, the pronoun and the headless relative express the same event-participant, and the predicate of the embedded clause is in the third-person:

---

[11] When discourse context leaves no room for ambiguity as to the syntactic roles of the third-person arguments of the verb in the embedded clause, object-centered relatives of both types (i.e. object-relatives and headless relatives) are possible:

i. *tuɬiltubuɬ ʔə* ***ti sqigʷəc tuqʼʷəxʷəd***

tu=ɬil–txʷ–buɬ ʔə ti sqigʷəc tu=qʼʷəxʷ–əd
PAST=give.food–ECS–1PL.OBJ PR DEF deer PAST=butchered–ICS
'he gave us the deer which he had butchered'

(Hukari 1977: 53)

ii. *taxʷčəɬəb sʔəɬəd ʔə* ***tiʔəʔ diʔəʔ stawixʷʔɬ tasčəbaʔəd tulʼʔal tudiʔ čaʔkʷ***

tu=ʔas–dxʷ–čəɬ–əb sʔəɬəd ʔə tiʔəʔ diʔəʔ stawixʷʔɬ
PAST=STAT–CTD–make–DSD food PR PROX here children
tu=ʔas–čəbaʔ–əd tulʼʔal tudiʔ čaʔkʷ
PAST=STAT–backpack–ICS PR over.there waterward
'she wanted to make food of the children she carried up from the water'

[DM Basket Ogress, line 73]

Even in such cases, object-centered relatives are unusual, the more common pattern being for the embedded verb to be used in the passive voice.

(23) a. *ʔəca* ***tiʔəʔ ləčalad tə sqʷəbayʔ***

ʔəca tiʔəʔ lə=čala–d tə sqʷəbayʔ
I PROX PROG=chased–ICS NSPEC dog
'the one who is chasing the dog is me'

b. *dibəł* ***ti ʔut'uc'utəb ʔə tiʔiɬ šəbad***

dibəł ti ʔu–t'uc'u–t–b ʔə tiʔiɬ šəbad
we SPEC PFV–shot–ICS–PASS PR.DIST enemy
'the ones who were shot by the enemy are us'

(Hess 1995: 99)

c. *gʷəl dəgʷi* ***kʷi ɬukʷədatəb dᶻixʷ***

gʷəl dəgʷi kʷi ɬu=kʷəda–t–b dᶻixʷ
then you REM IRR=held–ICS–PASS first
'and so the one who will be taken first is you'

[LA Basket Ogress, line 26]

These headless relative clauses are obligatorily subject-centred. When expression of the PATIENT or ENDPOINT of the event is the sentence predicate, the subject-phrase appears in the passive voice, as in (23b) and (c).

The syntactic properties of event-words in argument position in Lushootseed point quite clearly to their analysis as embedded predicates contained within a relative construction, something which is quite widely accepted as evidence of structural markedness in other contexts (for instance, in the discussion of property-concept verbs in languages with reduced classes of adjectives — e.g., Dixon 1982; Hengeveld, Rijkhoff and Siewierska 2004). Even though in Lushootseed there are no overt markers of this embedding such as special inflections or unambiguous complementizers, the added syntactic complexity of these embedded structures is manifest in the patterns of accessibility and voice restrictions discussed in the sections above. All of these properties of event-words in argument position point to their being significantly different (and more complex) constructions than a simple English noun phrase like *the boy*. Kinkade (1983) addresses this point by arguing, in effect, that substantive syntactic predicates like those discussed in Section 3.1 are evidence that a word like Lushootseed *sbiaw* 'coyote' in (4a) is the expression of the underlying semantic predicate 'be a coyote'.[12] Thus, according to Kinkade, all the translation-equivalents of nouns in languages like English are, in Salishan languages, the expressions of semantic predicates based on 'be'. If this is the case, then it must be true that not only are substantives predicative in constructions such as (4a), but they must also be predicates in sentences where they are syntactic arguments, as in (24):

[12] Kinkade's proposal is given a formal treatment in Jelinek and Demers (1994).

(24) *ʔuʔəɬəd tsi č'ač'as ʔə ti bəsqʷ*

ʔu–ʔəɬəd	tsi	č'ač'as	ʔə	ti	bəsqʷ
PFV–feed.on	SPEC:FEM	child	PR	D	crab

'the girl fed on crab'

(Hess 1995: 28, ex. 12b)

Crucial to this line of reasoning is the fact that in Lushootseed, as in most Salishan languages, nominal arguments are almost invariably introduced by determiners, as are headless relative clauses.[13] Thus, under Kinkade's analysis, each of the argument phrases in (24) would actually be the equivalent of a relative clause, the determiner in reality being a complementizer. According to Kinkade, Lawrence Nicodemus, a native speaker of Coeur D'Alene Salish with some linguistic training, regularly glosses argument phrases as relative clauses, as in:

Coeur D'Alene

(25) *x̌esiɬc'əʔ xʷe c'iʔ*

x̌es•iɬc'əʔ	xʷe	c'iʔ
good•flesh	DET	deer

'they are good to eat those which are deer'

(Nicodemus 1975, cited in Kinkade 1983: 34)

A better literal gloss for Kinkade's purposes might be 'the ones who are deer are good meat', the deictic *x□e* introducing a headless relative clause formed from *c'i□* 'be a deer'. Presumably, Nicodemus would also gloss the Lushootseed sentence in (24) as 'the one who is a girl fed on the one who is a crab'. The facts that subjects are gapped in subject- relative clauses and intransitive verbs show no overt agreement with third-person subjects lend a semblance of credibility to Kinkade's position in that if *tsi č'ač'as* in (24) were a subject- relative clause formed on a predicate 'be a child' with a zero subject, this is the form that it would be expected to take, a determiner followed by a bare intransitive predicate—cf. *ti ʔux̌ʷ* 'the one who goes' in (4a) (a similar point is made in Van Eijk and Hess 1986: 324–325).

Although Kinkade's interpretation of Nicodemus has had a certain intuitive appeal, it is difficult to know how seriously to take such considerations. What is needed before accepting such a radical claim — that all NPs are, in fact, syntactically relative clauses — is hard syntactic evidence, evidence which captures aspects of Salishan syntax and differentiates it from languages like English where such a position is clearly undesirable (although it has been argued for in the past, Bach 1968). So far none has been forthcoming, or at least none that cannot be handled in other ways such as a DP-analysis of noun phrases (Matthewson and Davis 1995; Beck 1997), which nonetheless maintains the noun–verb distinction. In the absence of such evidence, the more parsimonious analysis is to treat argument-phrases such as *tsi č'ač'as* in (24) as a simple

[13] In actual fact, proper names in Lushootseed often appear without a determiner. In addition, there are rare instances of bare common nouns in texts, although the conditions on this are not well understood. It should also be pointed out that nouns used as appositives, as predicate complements in negatives (see 35 below), and as complements in constructions with meanings like 'make an X' do not require a determiner, but do not have predicative 'be an X' readings.

substantive preceded by a determiner, a structurally less complex (and therefore unmarked) construction than that required for event-words in the same position, which are clearly relative clauses.

3.2.2 Oblique-centered constructions

Further evidence for the markedness of event-words in argument position, and the formal identity of these constructions with relative constructions, comes from the consideration of oblique-centered modifying and argument phrases. Since Lushootseed does not allow the relativization of oblique objects or adjuncts, it creates the structural equivalent of oblique- and adjunct- relative clauses through the formation of gerund- or participle-like constructions. Such constructions are used both as adnominal modifiers and as syntactic arguments, and have essentially the same internal syntax as matrix clauses in terms of the valency and transitivity of the embedded predicate. However, an important difference between the two clause types is that these non-finite clauses mark their subjects with the possessive series of subject-marking clitics, as shown in (26) for oblique-centered constructions formed with the proclitic *s*=, generally analyzed by Salishanists as a nominalizer:

(26) a. *x̌ʷul' čəd ɬuləʔuxʷtxʷ tiʔəʔ ɬadsʔəɬtxʷ*

x̌ʷul' čəd ɬu=lə–ʔux̌ʷ–txʷ tiʔəʔ ɬu=ad=s=ʔəɬ–txʷ
just 1SG.SUB IRR=PROG–go–ECS PROX I RR=2SG.PO=NM=eat–ECS
'I will just be taking [them] what you will feed [them] with'
(lit. 'I will just be taking them your future-feeding them')
(Hess 1998: 58, line 56)

b. *huyəxʷ tiʔiɬ dsyəhubtubicid, siʔab dsyaʔyaʔ*

huy=əxʷ tiʔiɬ d=s=yəhub–txʷ–bicid siʔab d–syaʔyaʔ
be.done=now DIST 1SG.PO=NM=recite–ECS–2SG.OBJ noble 1SG.PO-friend
'my telling to you is finished now, my noble friend'
(Hess 1995: 142, line 51)

In (26a) the subject of the non-finite clause *ɬadsʔəɬtxʷ* 'your future feeding him/her/them' (based on the transitive *ʔəɬtxʷ* 'feed someone with something') is expressed by the second-person singular possessive subject clitic, *ad*=. Similarly, in (26b) the subject of *syəhubtubicid* 'telling to you' is expressed by the first-person singular subject proclitic, *d*= (cf. the first-person matrix subject marker *čəd* in 20 above). The full set of possessive subject markers is given in Fig. 7:

	SG	PL
1	*d*=	*čəɬ*
2	*ad*=	*=ləp*
3	*=s*	

Fig. 7: Possessive subject-markers

This set is somewhat heterogeneous as it includes two proclitics, two enclitics, and a first-person plural clitic borrowed from the matrix subject paradigm shown in (7) above.

When the possessive subject is a full NP, a periphrastic construction with the preposition □*ə* is used (see 31 below).

The possessive subject paradigm is homophonous with the paradigm of affixes used to mark nominal possession:

(27) *d–sqʷəbayʔ* ‘my dog’
adˇ–sqʷəbayʔ ‘your dog’
sqʷəbayʔ–s ‘his/her/their dog’
sqʷəbayʔ čəɬ ‘our dog’
sqʷəbayʔ–ləp ‘your$_{PL}$ dog’
sqʷəbayʔ ʔə ti wiwˇsu ‘the children’s (*wiwˇsu*) dog’

However, the possessive subject markers differ from the possessive affixes in that they are mobile, and are obligatorily attached to the first element in the nominalized clause, whether or not this element is the sentence predicate, as in (28):

(28) *ʔa əwˇə sixʷ tiʔiɬ adsuhuy ti ƛ̓ˇubəstiləbsəxʷ ƛ̓ˇubəšəq*
ʔa əwˇə sixʷ tiʔiɬ ad=s=ʔu–huy
be.there PTCL PTCL DIST 2SG.PO=NM–PFV–be.done
ti ƛ̓ˇu=bə=s=tiləb=s=əxʷ ƛ̓ˇu=bə=šəq
SPEC HAB=ADD=NM–suddenly=3PO=now HAB=ADD=high
‘there is something you do to make it suddenly go high again’
(lit. ‘what you do [so that] it suddenly goes high again is there [i.e., exists]’)
(Hess 2006: 26, line 102)

In this example, the adverbial *ti ƛ̓ˇubəstiləbsəxʷ ƛ̓ˇubəšəq* ‘its habitually suddenly being high again’ contains an adverb *tiləb* ‘suddenly’ which precedes the clausal predicate *šəq* ‘be high’, and it is the adverb (rather than the clausal predicate) that bears both the possessive subject clitic and the nominalizing proclitic. Possessive affixes, on the other hand, are not mobile and remain affixed to the possessed, even in the presence of a pre-posed modifier (see the phrase *siʔab dsyaʔyaʔ* ‘my noble friend’ in 26b above, where the first-person possessive remains on *syaʔyaʔ* ‘friend’ rather than migrating to *siʔab* ‘noble’).[14]

As the example in (28) shows, it is not only the possessive subject clitics that are mobile, it is also the nominalizing proclitic itself. This property differentiates it from the homophonous (and certainly cognate) nominalizing prefix *s-* which forms a part of a great many nouns whose etymology is transparently that of a verbal radical plus this prefix, as well as a great many more where the etymology is no longer transparent. For many nouns formed with *s-*, the meaning of the derived form is fairly predictable: for intransitive verbs, the *s*-form refers to the subject of the verbal radical (e.g., *qˇaxʷ* ‘be frozen’ > *sqˇaxʷ* ‘ice’), while for bivalent verbs it refers to the object (*xʷiʔxʷiʔ* ‘hunt

[14] Hess (p.c., 2006) reports that for some older speakers the possessive affixes were optionally mobile, although there are no attestations of this pattern in the current corpus. Even if this were a frequent pattern, clitic migration is obligatory for possessive subject clitics, whereas at best it is only optional for possessive affixes.

something' > *sxʷiʔxʷiʔ* 'game'). However, it is also very common for *s*-forms to have unpredictable, lexicalized meanings (e.g., *x̌əɬ* 'be sick' > *sx̌əɬ* 'sickness', *x̌aʔx̌aʔ* 'be forbidden' > *sx̌aʔx̌aʔ* 'in-laws'). Even more significantly, the *s*- prefix—unlike the *s*= proclitic—is not mobile and can never separate from the radical to which it is attached:

(29) *laʔbəxʷ haʔɬ stalx̌əxʷ*

laʔb=əxʷ haʔɬ s–talx̌=əxʷ
really=now good NP–be.able=now
'now he is really a very capable one'

(Hess 2006: 40, line 461)

Note also that *s*-forms do not require an expression of a possessor, whereas nominalizations with the *s*= proclitic always appear with a possessive clitic expressing their subject.

In addition to the proclitic *s*=, Lushootseed also has a second proclitic, *dəxʷ*=, which forms non-finite clauses with essentially the same syntactic properties as those formed with *s*=, including the use of possessive subject clitics to express their subjects:

(30) a. *ləliʔəxʷ tiʔəʔ cəxʷuʔibəš*

ləliʔ=əxʷ tiʔəʔ d=dəxʷ=ʔu–ʔibəš
different=now PROX 1SG.PO=ADNM=PFV–travel
'where I am traveling is different now'

(Hess 2006: 27, line 128)

b. *ƛ̓'uləbəlx̌ʷ ʔal tiʔiɬ čad dəxʷʔaləp*

ƛ̓'u=lə=bəlx̌ʷ ʔal tiʔiɬ čad dəxʷ=ʔa=ləp
HAB=PROG=go.by PR DIST where ADNM=be.there=2PL.PO
'he goes by there where you guys come from'

(Hess 2006: 66, line 592)

(30a) shows the first-person singular proclitic *d*= marking the subject of the non-finite clause *cəxʷuʔibəš* 'where I travel'. The next example in (30b) contains a non-finite clause with a second-person plural subject, which is in turn contained within an prepositional phrase acting as a locative adverbial modifier. The distinction between *s*= and *dəxʷ*= is, roughly, that *s*= forms the equivalent of oblique-centered relative clauses, whereas *dəxʷ*= is used with adjunct-centered expressions referring (among other things) to locations, motives, and instruments.

When the subject of either type of non-finite clause is third-person, it shows the same patterns as the expression of the third-person possessor, using the possessive subject enclitic *=s* if there is no overt subject NP, otherwise making use of a periphrastic possessive construction:

(31) a. *x̌ʷul' p'aƛ̓'aƛ̓' tiʔiɬ sʔabyids tiʔiɬ č'ƛ̓'aʔ*

x̌ʷul' p'aƛ̓'aƛ̓' tiʔiɬ s=ʔab–yi–d=s tiʔiɬ č'ƛ̓'aʔ
just worthless DIST NM=extend–DAT–ICS=3PO DIST rock
'what he gives to that rock is simply worthless'

(Hess 1995: 148, line 32)

b. *tiʔiɬ tusʔukʷukʷ ʔə tə wiw'su*

tiʔiɬ	tu=s=ʔukʷukʷ	ʔə	tə	wiw'su
DIST	PAST=NM=play	PR	NSPEC	children

'what the children were playing with'

(Hess 1998: 89, line 299)

In (31a) the subject is realized with the third-person possessive subject marker, *=s*, while in (31b) the subject is an overt NP, *tə wiw'su* 'the children', and so the periphrastic possessive construction with *ʔə* is used. The same two patterns are also observed with *dəxʷ=*constructions:

(32) a. *ƛ'al' bədiɬ dəxʷʔa ʔə tiʔiɬ dəxʷʔəy'dubs ʔə tiʔiɬ sgʷəlub*

ƛ'al'	bə=diɬ	dəxʷ=ʔa	ʔə	tiʔiɬ	dəxʷ=ʔəy'–dxʷ–b=s	
also	ADD=FOC	ADNM=be.there	PR	DIST	ADNM=find–DC–PASS=3PO	
ʔə	tiʔiɬ	sgʷəlub				
PR	DIST	pheasant				

'it was the very same place where they had been found by Pheasant'

(Hess 1998: 85, line 187)

b. *ʔəsɬaɬlil tiʔiɬ ʔaciɬtalbixʷ dəxʷʔa ʔə tiʔacəc sbiaw*

ʔəs–ɬaɬlil	tiʔiɬ	ʔaciɬtalbixʷdəxʷ=ʔa		ʔə	tiʔacəc	sbiaw
STAT–live	DIST	person	ADNM=be.there	PR	UNQ	coyote

'people were living where Coyote was'

(Hess 1998: 91, line 1)

Again, here we see the use of the subject enclitic *=s* when there is no overt subject NP present (32a), and the periphrastic construction with *ʔə* used with an overt NP (32b).

A second function of the *s*=nominalizer is to form *sentential nominals* (Beck 2000), non-finite clauses whose reference is the event rather than a particular event-participant. Compare the non-finite clauses in (31) with those in (33):

(33) a. *tul't'aq't tiʔəʔ suʔəƛ' ʔə tiʔəʔ qʷuʔ*

tul'–t'aq't	tiʔəʔ	s=ʔu–ʔəƛ'	ʔə	tiʔəʔ	qʷuʔ
CNTRFG–waterward	PROX	NM=PFV–come	PR	PROX	water

'the coming of the water is waterward'

(Hess 1998: 69, line 108)

b. *ʔəsluud əlgʷəʔ tiʔiɬ suƛ'əladiʔs ʔal kʷədiʔ t'aq't*

ʔəs–lu–d	əlgʷəʔ	tiʔiɬ	s=ʔu–ƛ'əladiʔ=s		ʔal	kʷədiʔ
STAT–heard–ICS	PL	DIST	NM=PFV–make.noise=3PO		PR	REM:DMA
t'aq't						
waterward						

'they heard her making noise over there on shore'

(Hess 2006: 17, line 134)

The non-finite clauses in these examples refer to entire events — the coming of the water in (33a) and the making of a noise in (33b). In neither case is the reference of the non-finite clause an argument of the verb in the nominalized clause.

As it turns out, the interaction of the *s*=proclitic and words with a substantive meaning offers some evidence against the analysis of NPs as relative clauses and, as such, helps to establish the distinction between verbs and nouns. Consider the sentences in (34):

(34) a. *łustitčulbix^{w} čəx^{w}*

łu = s = titčulbix^{w} čəx^{w}

IRR=NM=small.animal 2SG.SUB

'you are the one who will be a small animal'

(Hess 2006: 8, line 136)

b. *huy, q^{w}iʔadəx^{w} tiʔəʔ skikəwič*

huy q^{w}iʔad = əx^{w} tiʔəʔ s = ki–kəwič

SCONJ call.out=now PROX NM=ATTN–hunchback

'then Little Hunchback calls out'

(lit. 'the one who is Little Hunchback calls out')

[AJ Basket Ogress, line 30]

Recall that Kinkade (1983) claims that all noun phrases in Salishan languages are in fact relative clauses, making an expression like *tiʔiɬ titčulbix^{w}* 'the small animal' in (34a) more literally 'the one who is a small animal' — that is, a subject- relative construction based on a monovalent predicate 'be a small animal'. Similarly, the form *skikəwič* in (34b) (based on the proper noun *kikəwič* 'Little Hunchback') would also seem to correspond to a subject- relative clause. If this were the case, however, then the occurrence of the proclitic *s*= with such words should, as it does with words expressing events, result in a sentential nominalization in which the subject is expressed as a possessor (meaning something along the lines of 'his being a small animal' or 'his being Little Hunchback'). Yet in the constructions in (34) the subject of the putative *s*=nominal is in fact not expressed at all; instead, such constructions seem to be interpreted as subject- relative clauses, which for event-words do not require the proclitic *s*=. Since *s*= is usually reserved for the 'relativization' of arguments that are not part of a predicate's core valency (i.e., not the subject or direct object), the obvious conclusion is that the subjects of substantive predicates are not in fact part of their core valency, which is consistent with the idea that substantives have a semantic valency of zero. This is fairly good evidence against the proposal that NPs are underlying relative clauses formed on 'be an X' type semantic predicates, given that such an analysis predicts that substantives should pattern in the same way as intransitive expressions of events and form full non-finite clauses when affixed with the *s*= proclitic.[15]

[15] An alternative analysis is that the /s/ here is not the nominalizing proclitic but rather the nominalizing prefix, *s*-, in which case a more accurate gloss of the forms in (34) might be something along the lines of 'the small-animal being' (34a) and 'the Little-Hunchback being' (34b). Even if this proves to be the case, the substance of the argument remains the same: the formation of sentential nominals is blocked for substantive predicates but allowed for words expressing events.

Although *s*= and *dəx*^{w}=clauses are not relative clauses *per se*, they represent the same type of structural markedness that headless relatives do — they are phrasal or clausal syntactic units and so count as being structurally complex. Further evidence for the markedness of these constructions of a different kind can adduced from the fact that they undergo a certain degree of *recategorization* (Bhat 1994) (also 'recategorialization' — Hopper and Thompson 1984) in that, by expressing their subjects as possessors, they take on some of the inflectional properties of the part of speech that is unmarked in the same syntactic role — that is, they become more like nouns (cf. Van Eijk and Hess 1986). Recategorization (and its counterpart, decategorization) falls under the heading of Contextual Markedness (Beck 2002: 23), and constitutes a clear example of what Hengeveld (1992a, 1992b) would classify as a 'further measure'. The fact that the morphosyntactic properties of event-words in Lushootseed in argument position become more like those of substantives seems to be strong evidence of the link between this syntactic role and the latter class of words — that is, evidence of the existence of a class of nouns in the Lushootseed lexicon.

3.2.3 Negative constructions

Once the existence of a lexical class distinction has been established by the examination of criterial syntactic environments, it is almost always the case that the distinction can also be shown to be present in other constructions as well. In Lushootseed, for instance, nouns and verbs can be shown to be clearly distinct in the context of negative constructions headed by the impersonal negative predicate *x*^{w}*iʔ* 'it is not, there is no'. This predicate is used to negate the existence or reality of an object or the realization of an event, and can take either a noun or a verb as its complement.[16] When the complement is a noun, the expression has the reading 'there is no' and the nominal complement appears with the subjunctive proclitic, *g*^{w}*ə*=:[17]

(35) a. *x*^{w}*iʔ g*^{w}*əstutubš*

x^{w}iʔ g^{w}ə = stu–tubš

NEG SBJ=ATTN–man

'there are no boys'

[LA Basket Ogress, line 119]

b. *x*^{w}*iʔ g*^{w}*əstabəx*^{w}

x^{w}iʔ g^{w}ə = stab = əx^{w}

NEG SBJ=what=now

'there is nothing (left)'

[ML Mink and Tutyika II, line 101]

---

[16] The negation of propositions such as 'X is not a boy' or 'X did not reach it' is carried out by different means, identical for nouns and verbs, involving the use of *x*^{w}*iʔ* as an adverb and the negative mood marker *lə*= (see Hess 1995: 94–95). The point being made here only concerns the subcategorization patterns of *x*^{w}*iʔ* used as syntactic predicate.

[17] This proclitic, one of the phrase-level inflectional proclitics discussed in Section 3.1, indicates that the phrase that contains it refers to a non-existent entity or a non-achieved eventuality.

When the complement is a verb, the subjunctive proclitic also appears, but the verbal predicate is obligatorily nominalized with the proclitic *s=*:

(36) a. *xʷiʔ uʔxʷ gʷəsłaʔ ʔə tiʔəʔ čaləs*

xʷiʔ uʔxʷ gʷə=s=łaʔ ʔə tiʔəʔ čaləs–s
NEG PTCL SBJ=NM=arrive PR ROX hand–3PO
'his hand still cannot reach it'
(lit. 'there is no his hand's reaching it')

(Hilbert and Hess 1977: 23)

b. *xʷiʔəxʷ gʷəsx̌aabs dxʷʔal słčil ʔə tsiʔəʔ bədaʔs*

xʷiʔ=əxʷ gʷə=s=x̌aab=s dxʷ–ʔal s=łčil ʔə tsiʔəʔ
NEG=now SBJ=NM=cry=3PO CNTRPT–at NM=arrive PR PROX:FEM
bədaʔ–s
offspring–3PO
'(the baby) isn't crying (even) when her daughter arrives'

[HM Star Child, line 48]

As with all of the other distinctions discussed in this section, this difference in the treatment of the two classes of words is categorical and requires reference in the syntax to a word-class designation that maps a set of syntactic behaviours onto specific items in the lexicon — in other words, a designation which constitutes a part-of-speech distinction. The fact that the semantic makeup of one of the classes corresponds almost exactly with the semantic category of substantives and the other contains those meanings belonging to the semantic category of events points us squarely to the conclusion that the distinction is one that any typologically-responsible analyst would characterize as one between nouns and verbs.

2.2 Unidirectional flexibility in the Salishan lexicon

As seen in the preceding sections, Lushootseed shows some flexibility as to the treatment of substantives versus event-words in its syntax, but this flexibility is unidirectional and only applies in predicate position. In argument position, event-words are either contained within headless relative clauses or are recategorized as non-finite argument phrases with some of the morphological properties of nouns (specifically, that they realize their subjects as possessors). The same type of recategorization applies when substantives and event-words are found in certain types of complementizing constructions such as negatives. In these cases, as in the case of non-substantive arguments, the syntax of Lushootseed makes reference to the lexical class of a word in determining its syntactic treatment: words that express substantive meanings appear in syntactic argument position without further measures, whereas words expressing events must be contained in some sort of embedded (headless relative or non-finite) clause when used as syntactic arguments. The fact that this syntactic distinction groups words into the semantic classes that it does (substantive versus event-word) shows quite clearly that Lushootseed makes a robust lexical class distinction between noun and verb.

Of course, this is not to say that the Lushootseed part-of-speech system is precisely the same as the traditional Indo-European system, nor that nouns and verbs behave in

Lushootseed exactly as they do in English or Latin. Like other Salishan languages, Lushootseed quite freely allows nouns and other non-verbal elements to serve as syntactic predicates without requiring the use of a copula, thereby neutralizing lexical class distinctions between verbs and nouns in predicate position. The situation can be summarized as in Fig. 8:

	PREDICATE	ARGUMENT
NOUN	*unmarked*	*unmarked*
VERB	*unmarked*	***marked***

Fig. 8: Unidirectional flexibility between nouns and verbs

Thus, the flexibility displayed by Lushootseed is particular only to one of the relevant criterial syntactic positions, but the noun–verb distinction can by no means be said to be absent from the Lushootseed lexicon or irrelevant to Lushootseed syntax. Rather, the distinction between verbs and nouns is not relevant to the behaviour of either class of word in predicate position — but it is relevant in argument position, as it is in languages like English and others that are said to have a 'rigid' distinction between noun and verb. This means, of course, that Lushootseed is like English in that it does distinguish between nouns and verbs as they are defined by Hengeveld (1992a, 1992b), but that it differs from English in that it shows flexibility with respect to the behaviour of nouns in predicate position. This is the type of unidirectional flexibility that Evans and Osada (2005) point to as a frequent source of claims for the absolute neutralization of the noun–verb distinction. As they correctly observe, such flexibility is a necessary but not a sufficient condition for the claim that a particular language does not distinguish the two lexical classes.

4. Omnipredicative and precategorial languages

The type of unidirectional flexibility shown by Lushootseed (and, to the best of my knowledge, other Salishan languages as well) is a typical case of what Evans and Osada (2005) refer to as the 'omnipredicative' pattern of putative noun–verb flexibility — that is, languages where all open-class lexical items are claimed to be semantic predicates with a minimum syntactic valency of one, those words with substantive meanings expressing predications of the type 'be an X'. As argued in the preceding sections, there is no positive empirical evidence for such a claim in Lushootseed, and some syntactic evidence against it. Taking the opposite point of view, that Lushootseed does have a noun–verb distinction and argument phrases headed by substantives are not syntactic predications (that is, they are not headless relative clauses), allows for a satisfactory account of the behaviour of open-class words in criterial and non-criterial syntactic positions without recourse to needless exoticisms in the syntax or the structure of the lexicon. Given that the facts in other languages where similar claims of omnipredicativity have been made (e.g., Nootkan — Swadesh 1939, Nahuatl — Launey 1994) look, to the non-specialist at any rate, to be substantially the same, it would seem that the case for this type of language representing a true case of bidirectional noun–verb flexibility does not stand up to careful scrutiny.

There is, of course, a second logically-possible type of language that neutralizes the syntactic distinction between words with substantive meanings and those that express events. In these languages, rather than substantive interpretations of 'be an X' semantic

predicates being forced by syntactic context, it is the substantive interpretations of event-words that are context-dependent. Languages that follow this pattern fall under Evans and Osada (2005)'s heading of precategorial languages. Rather than subdividing lexical items into classes based on their syntactic behaviours and distributions, precategorial languages organize their lexicon around roots that are not specified for any particular syntactic distribution, and thus do not conform to any of the definitions of parts of speech offered in (1) above. The meanings of these roots are often claimed to be 'vague' (Hengeveld et al. 2004; Hengeveld and Rijkhoff 2005) and to remain indeterminate between substantive or event-readings until they appear in a particular syntactic context. This situation is illustrated in the following by-now-familiar examples from Tongan:

Tongan

(37) a. *naʔe siʔi ʔae akó*
PST small ABS school:DEF
'the school was small'

b. *ʔi ʔene siʔí*
in 3SG:POSS small:DEF
'in his/her childhood'

c. *naʔe ako ʔae tamasiʔi siʔi iate au*
PST study ABS child little LOC 1SG
'the little child studied at my house'

d. *naʔe ako siʔi ʔae tamasiʔí*
PST study little ABS child:DEF
'the child studied little'

(Tchekoff 1981: 4, cited in Hengeveld 1992b: 66)

This data shows the root *siʔi* in a variety of contexts — as a syntactic predicate (37a), as the complement of a preposition (37b), as an adnominal modifier (37c), and as an adverbial (37d). These represent all four of the criterial syntactic contexts listed in (1) and so, given the absence of obvious morphosyntactic differences between the instances of *siʔi* in the various contexts, it is claimed that this root conforms to the definition of all four lexical classes.[18]

An obvious objection to this interpretation of the data is, of course, that *siʔi* does not mean the same thing in each of these four contexts (Croft 2000; Vonen 2000; Beck 2002). In (37a), (37c), and (37d), *siʔi* expresses a semantic predicate, something like 'be small' or 'be of reduced scale or intensity'; however, in (37b), *siʔi* expresses a more substantive concept, 'childhood' — that is, 'stage of human development during which a person is mentally and physically immature (and therefore small)'. While clearly not random, the exact semantic relationship between the two meanings of the root is not

[18] In fact, (37b) does not present a canonical example of *siʔi* in a criterial syntactic position for a noun — a more convincing example would give *siʔi* as the argument of a verb rather than as the complement of a preposition.

transparent or predictable, but rather is reminiscent of the relationship between homophonous pairs of English words such as *hammer*$_{N}$ vs. *hammer*$_{V}$ or *cook*$_{N}$ vs. *cook*$_{V}$, which are generally held to be distinct lexical items related by a process of conversion (Mel'čuk 2006). Conversion posits the existence in the lexicon of homophonous but semantically-related words whereby each has a particular meaning associated with its use in a particular syntactic role or roles.[19] The semantic relationship between members of a conversive pair is not arbitrary but at the same time is not entirely predictable and is established by linguistic convention, requiring the speaker to learn and enter into the mental lexicon the particular meaning of a root associated with its use in a particular semantic role. Evans and Osada (2005) make a similar observation for a large number of noun–verb pairs in Mundari and suggest that this pattern is evidence that precategorial languages are those that make extensive use of conversion in the lexicon, the distinct meanings of words like *si?i* in (37) in fact constituting different lexical items (see also Vonen 2000).

Hengeveld and Rijkhoff (2005) counter this argument by claiming that data like that in (37) are evidence that roots of this type have 'vague' meanings which become specified only once the root appears in a given syntactic context. Rijkhoff (2008) illustrates this idea using Fig. 9, which is based on the Samoan data in (38):

Samoan

(38) a. *'Ua lā le aso*
PERF sun ART day
'the sun is shining today' (lit. 'the day suns')

b. *'Ua mālosi le la*
PERF strong ART sun
'the sun is strong' (lit. 'the sun strongs')
(Mosel and Hovdhaugen 1992: 80, 73, 74, cited in Rijkhoff 2008: 729)

	A	B	C	D	E	Highlighted properties of *lā*:
Slot: head of Clause	+		+		+	A C E ⇒verbal meaning (*lā* 'be_sunny')
Slot: head of 'NP'		+		+	+	B D E ⇒nominal meaning (*lā* 'sun')
Slot: modifier of 'noun'		+	+	+		B C D ⇒adjectival meaning (*lā* 'sunny')

Fig. 9: Meaning components of Samoan *lā* (A B C D E) (Rijkhoff 2008: 731)

According to Rijkhoff's proposal, the meaning of the root *lā* consists of a set of components or features which comprise the sum total of the components of the

[19] It should be pointed out that the claim of conversion is not (as suggested by some common misnomers for 'conversion'—'zero conversion' and, worse, 'zero-derivation') a claim that there is some sort of class-changing zero affixation. Zero affixation requires an explicit formal contrast between a zero and a non-zero exponent of the same category, and it is precisely the nature of conversion that there is none. Furthermore, the existence of zero derivational elements is, under any viable theory of morphological zeroes, at best problematic, and more likely impossible (see Mel'čuk 2006: Chapter 9 and the references therein for further discussion).

meanings of its contextualized use. Speakers learn to associate certain subsets of the semantic components of *lā* with its appearance in particular semantic slots, giving rise to the more specific meanings that are lexified as different words in languages such as English.

In a sense it is certainly true, even under the conversion analysis, that the different context-bound meanings of precategorial roots share certain semantic elements and could, possibly, be shown to be extensions or subsets of a single, abstract schematic meaning; however, it is not clear how this impacts on the question of parts of speech. Hengeveld and Rijkhoff (2005) imply that a schematic relationship between the meanings of two signs with homophonous signifiers is sufficient for their treatment as the same lexical item. Nevertheless, the fact remains that the context-bound meanings of precategorial roots are quite specific, and that speakers must learn and memorize which specific sub-schematic meaning of the precategorial root (or which subset of its semantic components) is associated with which particular syntactic context.[20] Even for a carefully chosen example like *lā*, where the context-specific meanings of the root seem almost to be predictable (assuming that these meanings are exactly the same as their English glosses), it remains the case that speakers must learn and memorize the fact that, for instance, in the 'head of "slot"' *lā* expresses the meaning components B D E (= 'sun') and not A D E (say, 'sunshine').

The issue of predictability of the meanings of roots in particular semantic slots is discussed extensively by Evans and Osada (2005: 367–375) for another putative precategorial language, Mundari, under their heading of 'compositionality'. For Evans and Osada, a pair of words such as *lā* 'be sunny' and *lā* 'sun' can only be considered instances of the same lexeme if the difference in their meaning as it is manifest in two different syntactic slots is predictable as a function of that syntactic environment. Even a small shift such as 'sun' to 'be sunny' (that is, 'environment is illuminated brightly by the sun') seems not to be entirely predictable (e.g. why does the predicative use of *lā* not mean 'be a sun'?). However, the problem becomes even more acute for a root like the Tonga *si?i* in (37) which in its predicative and modificative function means 'small' and in its argument function means 'childhood' (i.e., 'stage of human development during which a person is mentally and physically immature'). In this case, the meaning of the precategorial root would have to include not only a component meaning 'small' but also all of the components that make up the meaning of the much more complex notion of 'childhood'. Furthermore, a speaker would have to learn (and store) the information that *only* the meaning component 'small' is associated with its use in predicative and modificative syntactic roles, and that the meaning 'childhood' (and not 'child' or 'short person' or any other conceivable recombination of the semantic component-set comprising the union of the components of 'small' and 'childhood') is associated with its use in syntactic argument roles.

Speakers may (or may not) be aware of the semantic relationship between the two meanings of the precategorial root, but the fact of the matter is that the speaker's

---

[20] This is not to say that there may not be systematic patterns in these associations; however, the best attempt to-date to come up with a classificatory scheme for a precategorial language (Tongan — Broschart 1997) reveals a complex system in which membership in one of 24 lexical classes (as defined by these patterns) appears to be to a certain extent non-arbitrary but is by no means predictable. Acquisition of such a system may be facilitated by the semantics of roots, but class membership — that is, which semantic relationship holds between the predicative and substantive interpretations of the word — must nevertheless be learned (and stored in the lexicon) on an item-by-item basis.

knowledge of *si?i* must include a learned pairing between a particular meaning (sub-schematic or not) and a syntactic distribution. And this pairing of meaning and unmarked distribution is what is generally understood by 'part-of-speech'. The abstract schemas or vague meanings underlying the networks of related words may constitute part of the lexicon of a precategorial language (as may the schemas shared by non-homophonous forms related by overt derivation and other word-formation processes); however, it is not the precategorial roots that meet the definitions of lexical classes shown in (1), it is the context-bound uses of the roots. Like most approaches to parts of speech, the definitions in (1) are based on the premise that parts of speech define high-level taxonomic groupings of lexical items (i.e., lexical signs that pair a signifier and a particular signified) according to their (unmarked) syntactic distribution. For a language like Tongan, such definitions fit quite naturally if we recognize that the term 'lexical item' applies separately to each member of a pair like *si?i* 'be small'/*si?i* 'childhood', rather than to an under-specified, abstract root *si?i*, which seems more than anything else to define the semantic domain of the word-family. Like the analysis of omnipredicative languages suggested above, this approach effectively removes precategorial languages from the putative group of languages that show bidirectional noun–verb flexibility, and avoids treating the lexicon in such languages as being overly exotic, instead showing them to be extreme examples of a lexicon built on the cross-linguistically well-attested process of lexical conversion.

Seen from this perspective, the type of noun–verb flexibility manifested by omnipredicative languages and that seen in precategorial languages represent markedly different phenomena. In the former case, words with specific meanings fall into clear categories based on their morphosyntactic behaviour, but rather than showing bidirectional patterns of markedness in both of the criterial syntactic roles for nouns and verbs, the class of verbs shows itself to be marked in the role of syntactic argument, while both nouns and verbs are unmarked syntactic predicates. In precategorial languages, homophonous lexical items appear in various syntactic roles, but have divergent (albeit not always unrelated) meanings associated with their different uses. The homophonous lexical items undoubtedly form related sets and so may be linked to an abstract precategorial schema, but the sets are in a sense 'pre-lexical', representing a more abstract level of the lexicon than has traditionally been of interest to syntacticians and lexicographers. Irrespective of the internal structure of these sets, it is the individual members (that is, the distributionally-specified meaning–form pairs) that constitute the genuine parts of speech in such languages, making the precategorial language — like the omnipredicative language — a false case of noun–verb flexibility.

Thus, both possible types of noun–verb flexibility proposed by Evans and Osada (2005) seem (as they suggest) not to stand up to careful consideration. This is an unsurprising result, given the widely-held position among typologists that the noun–verb distinction is one of the best candidates we have for a genuine universal of human language (Croft 2003). Naturally, this raises a question with respect to the status of the typology of parts-of-speech systems in Fig. 1 and the implicational hierarchy in (2), given that both predict, or at any rate allow for, the existence of languages that do not distinguish verbs and nouns. While this may not be a terribly serious shortcoming of the typology, as it could simply be stipulated that all languages make at least the first-order distinction on the hierarchy and the Type 1 language could then be removed from the typology, the fact that we need to make such a stipulation at all seems significant, and lays the groundwork for future investigation into the origins and motivations of this restriction on the logically-possible range of variation in human languages.

A second, and perhaps more serious, objection to the proposed typology that seems to fall out from the analysis of the data presented here has to do with the relative rankings of the different word-classes in the hierarchy, in particular with respect to the rankings of noun and verb. Even if it is the case that no language fails to distinguish between nouns and verbs, the way that the hierarchy is presently constructed characterizes languages that distinguish only two lexical classes as making a two-way distinction between verbs and everything else. However, the distributional evidence from both Salishan and Tongan seem to point us in the opposite direction: that languages with only two lexical classes distinguish between nouns and everything else. In Salishan, this is manifest in the pattern of distributional markedness. Nouns are syntactically privileged in that they have a syntactic role (syntactic argument) that is not open without further measures to other lexical classes, effectively subdividing the lexicon between words which are marked and unmarked syntactic arguments, rather than between words that are marked and unmarked syntactic predicates, as predicted by the typology in Fig. 1.[21] In the Tongan case, assuming that the pattern shown by *si?i* is typical of all precategorial roots in the language, the two interpretations of the root are divided along similar lines — the substantive interpretation is restricted to 'nominal' syntactic roles and the predicative interpretation is found elsewhere, once again dividing the lexicon between words (or interpretations of precategorial roots) which are without-further-measures syntactic arguments and words which are found without further measures in the other criterial syntactic environments. Thus, it seems that the cross-linguistically attested patterns of flexibility in lexical classes favour a typology, like those argued for in Dixon (1982) and Beck (2002), that allows for the flexible grouping of nouns against verbs, adjectives, and adverbs rather than verbs against a potentially-conflated class of nouns, adjectives, and adverbs. From a semantic perspective, this makes a great deal of sense, given that verbs, adjectives, and adverbs are expressions of semantic predicates and share the property of having non-zero syntactic valency, as opposed to nouns which in Langacker's (1987) terms are 'conceptually autonomous' and generally have a syntactic (and semantic) valency of zero. If this pattern turns out to be cross-linguistically robust, it constitutes an important finding, as it offers us an example of the influence of the semantic structure (as opposed to the content) of meanings on the organization of their expressions in lexicon.

A final implication of this study for the approach to parts-of-speech typology being discussed here concerns the issue of the directionality of lexical flexibility. Although the principle of bidirectionality was proposed by Evans and Osada (2005) as a test for determining whether or not a part-of-speech distinction is truly absent in a language, it also sheds light on an important potential difference in types of lexical flexibility, highlighted by the contrast between Fig. 3 and Fig. 4 above, repeated here for convenience in Fig. 10:

[21] It should be pointed out here that, in addition to verbs, Lushootseed has a number of smaller, closed classes of word — one of which, adverbs, contains what are often thought of as 'contentive' (as opposed to 'grammatical') meanings. The fact that Lushootseed has adverbs but not adjectives is also somewhat problematic from the point of view of the typology in Fig. 1, although this is mitigated by the fact that adverbs are a closed class of words.

	BIDIRECTIONAL		UNIDIRECTIONAL	
	ROLE A	ROLE B	ROLE A	ROLE B
CLASS X	*unmarked*	*unmarked*	*unmarked*	*unmarked*
CLASS Y	*unmarked*	*unmarked*	*unmarked*	***marked***

Fig. 10: Types of flexibility

While flexibility between lexical classes has often been assumed to entail bidirectionality, languages like Lushootseed provide us with a clear example of a different, unidirectional type of flexibility. Unidirectional flexibility does not equate with the absence of a lexical class distinction, but it does correspond to the neutralization of that distinction in one (or more) of the criterial syntactic positions. This type of neutralization is a relative commonplace across languages, and seems like a good candidate for inclusion as a parameter for a comprehensive typology of parts of speech systems. The patterns shown by unidirectional systems and the ways in which they parallel and depart from the patterns observed for bidirectionally-flexible systems seem sure to inform parts-of-speech typology and will advance the cause of understanding the parameters of variation open to human languages.

References

Bach, E. (1968). 'Nouns and noun phrases', in E. Bach and R. T. Harms (eds.), *Universals in Linguistic Theory*. New York: Holt, Rinehart, and Winston, 91–124.

Bates, D. (2002). 'Narrative functions of past tense marking in a Lushootseed text', in C. Gillon, N. Sawaii and R. Wojdak (eds.), *Proceedings of the 37th International Conference on Salishan and Neighbouring Languages*. Vancouver: UBCWPL, 17–34.

Bates, D., Hess, T. M. and Hilbert, V. (1994). *Lushootseed dictionary.* Seattle: Washington University Press.

Beck, D. (1997). 'Theme, Rheme, and Communicative Structure in Lushootseed and Bella Coola', in L. Wanner (ed.), *Recent trends in Meaning-Text Theory*. Amsterdam: Benjamins, 93–135.

--- (2000). 'Nominalization as complementation in Bella Coola and Lushootseed', in K. Horie (ed.), *Complementation: Cognitive and functional perspectives*. Amsterdam: Benjamins, 121-147.

--- (2002). *The typology of parts of speech systems: The markedness of adjectives.* New York: Routledge.

--- (in progress). 'Lushootseed morphosyntax', University of Alberta.

Beck, D. and Hess, T. M. (n.d.). 'Lushootseed *syəyəhub* with interlinear analyses', University of Alberta.

Bhat, D. N. S. (1994). *The adjectival category: Criteria for differentiation and identification.* Amsterdam: Benjamins.

Broschart, J. (1991). 'Noun, verb, and participation (a typology of the Noun/Verb distinction)', in H. Seiler and W. Premper (eds.), *Partizipation: Das sprachliche Erfassen von Sachverhalten*. Tübingen: Narr, 65–137.

--- (1997). 'Why Tongan does it differently: Categorial distinctions in a language without nouns and verbs', *Linguistic Typology* 1: 123–165.

Burton, S. (1997). 'Past tense on nouns as death, destruction, and loss', *NELS* 27: 65–77.
Croft, W. (1991). *Syntactic categories and grammatical relations: The cognitive organization of information.* Chicago: University of Chicago Press.
--- (2000). 'Parts of speech as language universals and as language-particular categories', in P. M. Vogel and B. Comrie (eds.), *Approaches to the typology of word classes*. Berlin: Mouton, 65–102.
--- (2003). *Typology and universals.* Cambridge: Cambridge University Press.
--- (2005). 'Word classes, parts of speech, and syntactic argumentation', *Linguistic Typology* 9: 431–441.
Davis, H. and Matthewson, L. (1998). 'Definiteness, finitieness, and the event/entity distinction', *NELS* 28: 95–109.
--- (1999). 'On the functional determination of lexical categories', *Revue québécoise de linguistique* 27: 27–69.
Demirdache, H. and Matthewson, L. (1995). 'On the universality of syntactic categories', *NELS* 25: 79–94.
Dik, S. C. (1997). *The Theory of Functional Grammar.* Berlin and New York: Mouton de Gruyter.
Dixon, R. M. W. (1982). *Where have all the adjectives gone?* Berlin: Mouton de Gruyter.
Evans, N. and Osada, T. (2005). 'Mundari: The myth of a language without word classes', *Linguistic Typology* 9: 351–390.
Hengeveld, K. (1992a). *Non-verbal predication: Theory, typology, diachrony.* Berlin: Mouton de Gruyter.
--- (1992b). 'Parts of speech', in M. Fortescue, P. Harder and L. Kristofferson (eds.), *Layered structure and reference in a functional perspective*. Amsterdam: Benjamins, 29–56.
Hengeveld, K. and Rijkhoff, J. (2005). 'Mundari as a flexible language', *Linguistic Typology* 9: 406–431.
Hengeveld, K., Rijkhoff, J. and Siewierska, A. (2004). 'Parts-of-speech systems and word order', *Journal of Linguistics* 40: 527–570.
Hess, T. M. (1967). *Snohomish grammatical structure.* Seattle: University of Washington.
--- (1995). *Lushootseed reader with introductory grammar, Volume I.* Missoula: University of Montana Occasional Papers in Linguistics.
--- (1998). *Lushootseed reader, Volume II.* Missoula: University of Montana Occasional Papers in Linguistics.
--- (2006). *Lushootseed reader, Volume III.* Missoula: University of Montana Occasional Papers in Linguistics.
Hess, T. M. and Hilbert, V. (1976). *Lushootseed: An introduction, Books 1 and 2.* Seattle: American Indian Studies, University of Washington.
Hilbert, V. T. and Hess, T. M. (1977). 'Lushootseed', in B. F. Carlson (ed.), *Northwest Coast texts: Stealing light*. Chicago: University of Chicago Press, 4–32.
Hopper, P. and Thompson, S. A. (1984). 'The discourse basis for lexical categories in universal grammar', *Language* 60: 703–752.
Hukari, T. E. (1977). 'A comparison of attributive clause constructions in two Coast Salish languages.', *Glossa* 11: 48–73.

Jacobsen, W. H. (1979). 'Noun and verb in Nootkan', in B. Efrat (ed.), *The Victoria Conference on Northwestern Languages, Heritage Record #4*. Victoria, BC: British Columbia Provincial Museum, 83–153.
Jelinek, E. and Demers, R. A. (1994). 'Predicates and pronominal arguments in Straits Salish', *Language* 70: 697–736.
Jesperson, O. (1965). *Philosophy of Grammar.* London: W. W. Norton.
Keenan, E. and Comrie, B. (1977). 'Noun phrase accessibility and universal grammar', *Linguistic Inquiry* 8: 63–99.
Kinkade, M. D. (1983). 'Salishan evidence against the universality of "noun" and "verb"', *Lingua* 60: 25–40.
Kroeber, P. D. (1999). *The Salishan language family: Reconstructing syntax.* Lincoln, NB: University of Nebraska Press.
Kuipers, A. H. (1968). 'The categories verb-noun and transitive-intransitive in English and Squamish', *Lingua* 21: 620–626.
Langacker, R. W. (1987). *Foundations of cognitive grammar.* Stanford, CA: Stanford University Press.
Launey, M. (1994). *Une grammaire omniprédicative.* Paris: CNRS Editions.
Matthewson, L. (2002). 'Tense in St'at'imcets and in Universal Grammar', in C. Gillon, N. Sawaii and R. Wojdak (eds.), *Proceedings of the 37th International Conference on Salishan and Neighbouring Languages*. Vancouver: UBCWPL, 233–260.
--- (2005). 'On the absence of tense on determiners', *Lingua* 115: 1697–1735.
Matthewson, L. and Davis, H. (1995). 'The structure of DP in St'at'imcets (Lillooet Salish)', *Papers for the 30th International Conference on Salish and Neighbouring Languages*: 55–68.
Mel'čuk, I. A. (2006). *Aspects of the theory of morphology.* Berlin: Mouton de Gruyter.
Mosel, U. and Hovdhaugen, E. 1992. *Samoan reference grammar*. Oslo: Universitetsforlaget AS.
Nicodemus, L. (1975). *Snchitsúumshtn: The Coeur D'Alene language: A modern course.* Plummer, ID: Coeur D'Alene Tribe.
Rijkhoff, J. (2008). 'On flexible and rigid nouns', *Studies in Language* 32: 727–752.
Sasse, H.-J. (1993). 'Syntactic categories and subcategories', in J. Jacobs, A. von Stechow, W. Sternefeld and T. Vennemann (eds.), *Syntax: An international handbook of contemporary research, volume 1*. Berlin: de Gruyter, 646–686.
Swadesh, M. (1939). 'Nootka internal syntax', *International Journal of American Linguistics* 9: 77–102.
Tchekoff, C. (1981). *Simple sentences in Tongan.* Canberra: Australian National University.
Tesnière, L. (1934). 'Comment construire une syntaxe.', *Bulletin de la Faculté des lettres, Université de Strassbourg* 7: 219–229.
--- (1959). *Éléments de syntaxe structurale.* Paris: Klincksieck.
Van Eijk, J. and Hess, T. M. (1986). 'Noun and verb in Salishan', *Lingua* 69: 319–31.
Vonen, A. M. (2000). ' Polynesian multifunctionality and the ambitions of linguistic description', in P. M. Vogel and B. Comrie (eds.), *Approaches to the typology of word classes*. Berlin: Mouton, 479–487.
Wiltschko, M. (2003). 'On the interpretability of Tense on D and its consequences for Case Theory', *Lingua* 113: 659–696.

Lexical categories in Gooniyandi, Kimberley, Western Australia

William B. McGregor

1. Introduction

1.1 Background[1]

Received wisdom has it that the standard average Australian Aboriginal language is characterised by around ten distinct, and virtually disjoint, lexical classes: nouns, adjectives, verbs, pronouns, adverbials, locational qualifiers, temporal qualifiers, particles, interjections, and ideophones (Dixon 1980: 271).[2] These classes can be grouped into three major sets according to inflectional criteria:[3] nominals, verbs, and others (Dixon 2002: 66, 71; Blake 1987: 2-3). Thus (it is claimed) nominals take case inflections, and include nouns, pronominals, determiners, adjectives, locational qualifiers, and temporal qualifiers; verbs take tense-mood-aspect inflections, and include simple verbs, and (in some languages) coverbs (also called preverbs and uninflecting verbs) and/or adverbials; and the other classes admit no inflection, and typically include particles, interjections, ideophones, conjunctions, and (in some languages) adverbs. Derivational morphology is required for category changes: a nominal can take tense-mood-aspect inflections only if it is derived by a verbal derivational morpheme; a verb can take case-markers and other nominal morphology only if it is derived by a nominal derivational morpheme (Dixon 1980: 271, 2002: 75).

As remarked in the previous paragraph, in the typical Australian language the category of nominal includes both nouns and adjectives, which show identical inflectional potential. According to some accounts, nouns and adjectives represent two distinct lexical subclasses of the supercategory of nominals (Dixon 1980: 271, 2002: 66), distinguishable on conceptual grounds (Dixon 1980: 274-275, 2002: 68). Specifically, nouns are those lexemes that denote entities, whereas adjectives denote qualities. In a relatively few languages nouns and adjectives do show distinct morphological behaviour; for instance, in Unggumi (Worrorran, Kimberley), adjectives can be distinguished by virtue of the fact that they admit affixed class markers of all types, whereas nouns admit more restricted sets of affixed class marker (usually just one set).[4]

---

1 I am grateful to the editors and three anonymous referees for their extensive commentary on a previous draft of this paper, including corrections of a number of inaccurate statements concerning the functional discourse grammar approach to parts-of-speech. The usual disclaimers apply.

2 This listing of parts-of-speech is based on Pama-Nyungan languages, primarily from the eastern side of the continent, the languages that well known Australianists such as R.M.W. Dixon and Barry Blake are familiar with.

3 Dixon (1980: 271) avers that "[e]ach word class has its own set of inflections", though the subsequent discussion is at variance with this claim. In practice, Australianists usually employ a rag-bag of *ad hoc* criteria for identifying the parts-of-speech in particular languages, including morphological, syntactic, and notional-conceptual.

4 The class of adjectives so defined is very small, and includes only a few of the items that might be expected (on the basis of their apparent meaning) to fall into the class.

Granted the above remarks, it would seem reasonable to guess that the standard average Australian Aboriginal language is fairly rigid in terms of its parts-of-speech categories: one expects a one-to-one correspondence between the parts-of-speech and the grammatical functions they may serve. This would presumably be the case under the assumption of the maximal ten parts-of-speech system, in which, if a word of any part of speech serves in a role other than the one prototypically associated with it, a derivational morpheme would need to be employed to register the unusual grammatical function. The major uncertainty depends on whether or not nouns and adjectives (and to a lesser extent, in a smallish set of languages, verbs and adverbs) represent separate and viable parts-of-speech. If they don't, nominals would represent a flexible category of words that could function both as head and modifier in referring expressions.

1.2 Problems with the received construal of parts-of-speech

In this paper I demonstrate that the received view as outlined above runs into serious difficulties in at least one Aboriginal language, Gooniyandi (Bunuban, Kimberley, Western Australia). In particular, the simple morphological criteria identified in §1.1 do not carve out viable parts-of-speech categories or superclasses of categories. To begin with, inflection is highly restricted in the language, and is not a property of either words that are evidently verbs or words that show the hallmarks of nominals; indeed, it is only words of some closed classes that admit any inflection: pronominals, a few non-manner adverbials, and a set of bound morphemes I refer to as classifiers (McGregor 1990: 194) — i.e. mostly classes of items that we would expect from the received view to admit no inflections.

One solution to this problem would be to revise, perhaps by weakening, the standard morphological criteria. For instance, one could lift the requirements of inflection, and define the parts-of-speech in terms of the bound morphemes (mostly enclitics, but also a few derivational morphemes) that may be attached to a given lexical item. Unfortunately, this revision does not yield a viable categorisation of the Gooniyandi lexicon. For, as it turns out, there are few absolute restrictions on the types of bound morpheme that can be attached to particular lexemes. Thus, a lexeme such as *yoowooloo* 'man' can take a range of bound morphemes that are typically associated with nominals (case and number marking postpositions), as well as bound morphemes generally associated with verbs (see example (12) below).

One can, however, arrive at an intuitively plausible parts-of-speech grouping by defining verbs as bound lexemes, nominals as free lexemes that admit in principle any encliticised postposition, and adverbials as free lexemes that may host at best restricted ranges of postpositions, though never number marking postpositions. (Other parts-of-speech can be given similar characterisations in terms of their morphological potentials.) The problem with this approach is that the selection of morphological characteristics has no obvious theoretical motivation; the main points in its favour are that it results in a categorisation that accords well with intuition, and that it turns out to be descriptively useful. The extent to which one sees this as a serious concern will depend on the extent to which one accepts and abides by the Boasian tenet that grammatical descriptions should be first and foremost founded on facts of the particular language, to the exclusion of other languages and theoretical concerns.

McGregor (1990: 140-141) puts forward another solution, one that appears less arbitrary, and is in general terms in line with the Functional Grammar (FG) and

Functional Discourse Grammar (FDG) approaches (e.g. Hengeveld 1990, 1992; Hengeveld et al. 2004; Rijkhoff 2007; van Lier 2008, van Lier 2009; Hengeveld and McKenzie 2008). This approach is based on the grammatical relations that lexical items may serve; see §2.1 for details. This scheme results in a system of parts-of-speech comprising largely disjoint classes. The system shows, however, a fair amount of flexibility in the sense that lexical words of a given part-of-speech are rarely restricted to a single grammatical relation, but rather typically have the potential of discharging a range of different relations. Thus (putting things for now in FG terms — reformulations will be given below in accordance with my own theoretical approach) nominals can function as heads of predicate phrases and heads or modifiers of referential phrases, and adverbials as modifiers in referential phrases and heads of predicate phrases.

1.3 Aims and organisation of paper

I have three primary aims in this paper. First, I want to show that my 1990 approach to the definition of parts-of-speech in Gooniyandi can be refined and extended to encompass a wider range of parts-of-speech than the three it was designed for. In doing this, I invoke a revised and elaborated scheme of grammatical relations identified in Semiotic Grammar (SG) (McGregor 1997). Identification of the full extent of the parts-of-speech system in a language is a necessary though insufficient step in the achievement of the second aim, which is to identify the extent of flexibility of the parts-of-speech. Third, I argue that the semantics of members of flexible parts-of-speech is quite abstract and vague, and that the meanings of particular instances of use of words belonging to the flexible parts-of-speech are predictable from their coded meanings, the meanings of the grammatical relations and constructions they enter into, and pragmatic inferences. This argues against the notion that members of flexible parts-of-speech should be treated as distinct, though homophonous, lexemes.

The paper is structured as follows. I begin in §2.1 by outlining the characterisation of the Gooniyandi parts-of-speech system, starting with the system proposed in McGregor (1990); the motivation for beginning with this characterisation is that it provides an easy path into some of the difficulties in defining parts-of-speech categories in the language. Section 2.2 then shows how some of these problems can be resolved by a reformulation in accordance with the SG model, which is outlined briefly. Following this, §3 briefly discusses some issues in semantics, and proposes that Gooniyandi lexemes are prototypically vague in terms of their coded meanings. Section 4 winds up the paper with a summary and concluding remarks.

2. Parts-of-speech categories in Gooniyandi

2.1 McGregor's (1990) characterisation of Gooniyandi parts-of-speech

McGregor (1990: 140-141) (henceforth FGG — an acronym for *A functional grammar of Gooniyandi*) provides the following characterisation of the three major parts-of-speech in Gooniyandi:

VERBALS are those lexemes which must realise the role of Event in a verb phrase (VP);
NOMINALS are those which must either realise some function in a noun phrase (NP), or (less frequently) the Event in a VP; and
ADVERBIALS are those lexemes which may realise functions in clauses, NPs, postpositional phrases (PPs) or VPs.

To make sense of these characterisations it is necessary to lay out some background information on the phrasal units and grammatical functions invoked. Let us begin with the three types of phrasal unit.

An NP is a grammatical unit constituted by a coherent and unified sequence of words that refers to an entity of some sort (e.g. concrete or abstract), and that may realise (among other things) a clausal role such as Actor or Undergoer (see below). As argued in FGG: 253ff, the NP in Gooniyandi can be described structurally-functionally in terms of a set of grammatical roles that appear in the strict sequence (Deictic)-(Quantifier)-(Classifier)-Entity-(Qualifier). These roles are typically discharged by word-sized expressions, though sometimes larger syntagms are deployed (including NPs and PPs); the expression sets for the roles overlap somewhat, but are significantly different. As this description indicates, all roles except for the Entity are optional; the Entity role is inherent, though it can be ellipsed if it presents predictable information. (1) and (2) are illustrative examples of NPs; the first line shows the phrasal roles realised by the component words. PPs are grammatical units constituted by an NP in collocation with a postposition, which may serve either as a case-marker or a number marker. As illustrated by (3) and (4), postpositions are phrase-level enclitics that occur one per phrase; they may be attached to any word of the phrase, though usually they go on the most prominent or focal one, roughly the one that conveys the most significant information.[5]

(1)	Deictic	Quantifier	Entity	Qualifier
	ngoorroo	*garndiwirri*	*yoowooloo*	*gima-ngarna*
	that	two	man	bush-DW
	'those two bushmen'			

5 Example sentences are laid out in accordance with the Leipzig Glossing Rules, using the recommended abbreviations. For the reader's convenience a full list of abbreviations is provided: ABL — ablative; ACC — accusative; ALL — allative; CRD — cardinal; CHAR — characteristic; CNT — continuous; COM — comitative; DAT — dative; DU — dual; DW — dweller of a niche; ERG — ergative; EXC — exclusive; FUT — future; GD — good at; INC — inclusive; INF — infinitive; IT — iterative; LOC — locative; MD — mode; MNR — manner; NOM — nominative; OBL — oblique; PER — perlative; PL — plural; PROG — progressive; PRS — present; RDP — reduplicant; REP — repeated ('again'); SG — singular; and SUB — subjunctive. In addition to these abbreviations, I employ the following conventions: numerals 1-3 for the three persons; a full stop (.) to separate abbreviations in glosses; a colon to separate the words of multiword glosses for a single morpheme; + to indicate a particular type of morpheme boundary found within a part of the verb; a slash (/) in the text line to indicate the end of an intonation contour (only in textual examples), and in the gloss line to indicate portmanteaus which cannot be segmented into separate morphemes; lastly, classifiers in the verb are not provided glosses, but are represented in a citation form given in capitals (following a convention adopted in other works on the language). The following abbreviations are used in the text: FG — functional grammar; FDG — functional discourse grammar; FGG — *A functional grammar of Gooniyandi* (McGregor 1990); NP — noun phrase; PP — postpositional phrase; SG — semiotic grammar; SoA — State-of-Affairs; and VP — verb phrase.

(2) Quantifier Classifier Entity
garndiwangoorroo *thiwa* *goornboo*
many red woman
'many white women', 'many women of European descent'

(3) *ngoorroo-ngga* *ngarloordoo* *yoowooloo*
that-ERG three man
'by those three men'

(4) *ngoorroo-yarndi* *yoowooloo*
that-PL man
'those men'

FGG distinguishes two primary types of clause in Gooniyandi, relational clauses and situation clauses (FGG: 293). Relational clauses establish relations among phenomena (widely construed to include entities, qualities, locations, and so forth), and typically consist of an NP in apposition to another NP, PP, or adverbial, the two being related by a relation such as elaboration (where one unit states the other in a different way), extension (in which an element adds to the other), or enhancement (in which one element provides circumstantial information concerning the other, e.g. where it is located in time or space) (see also McGregor 1996b). (5) is an example of a relational clause in which the relation is identification: the referent of the first NP 'this place' is identified with the referent of the second NP 'our (place)', which is an elliptical NP consisting of just the oblique pronominal *ngirrangi*, the Entity nominal *riwi* 'place' being ellipsed. Relational clauses have effectively no temporal structure, and time is at best relevant as a locus for the relation, a temporal domain within which it obtained.

(5) *ngirndaji* *riwi* *ngirrangi*
this place 1PL.EXC.OBL[6]
'This place is ours.'

Situation clauses, by contrast, construe phenomena that unfold over time, including states, events, happenings, processes, activities, and so on, and which involve one or more entities that are crucially engaged in them. Correspondingly, the clause is made up of linguistic elements that serve to specify the ongoing phenomenon, and linguistic elements that serve to specify the entities engaged in them. Thus situation clauses are constituted by a State-of-Affairs (SoA) role (specifying the ongoing phenomenon), and one or more participant roles, roughly corresponding to what are usually dubbed argument roles. (6) and (7) provide illustration, with two simple situation clauses, one intransitive, the other transitive (there are other construction types in the language). Again the first line indicates the grammatical roles; the role labels are explained in

6 Gooniyandi distinguishes an unusual type of inclusive-exclusive contrast in the first person non-singular, a contrast between including and excluding addressees (i.e. the addressee plus one or more others) rather than a single addressee. The details of this system need not detain us here; see McGregor (1996c: 159-173) for discussion.

FGG, (McGregor 1997), and in various other publications (e.g. McGregor 2006), and are irrelevant to the present exposition.

(6) Actor/Medium SoA
nganyi *ward-ngi*
1SG.CRD go-1SG.NOM/I
'I went.'

(7) Actor/Agent Undergoer/Medium SoA
nganyi-ngga *wayandi* *jard-li*
1SG.CRD-ERG fire put:flame:to-3SG.ACC/1SG.NOM/DI
'I lit a fire.'

Participant roles are realised by either NPs or PPs. The SoA role is by contrast realised by a VP, which is characterised by an inherent experiential role, the Event,[7] and an optional element in an attributing relation to this Event, typically specifying a quality of its performance (e.g. speed, or manner). The VP in this conception does not include direct and indirect object NPs or other elements such as locationals, directionals, and so on, i.e. the VP is not construed as consisting of everything bar the "subject" NP, as it is in many frameworks, including formal ones. (For further discussion of this phrasal category see McGregor 1996a.)

With this background in place, let us return briefly to the three primary parts-of-speech to see how well the above characterisations fare. Verbals are defined in FGG by their restriction to serving the Event role in VPs. Items such as *ward-* 'go' and *jard-* 'put flame to' can occur only in VPs, either finite ones as in (6) and (7), or non-finite ones, as in (8) and (9).

(8) *ngirrinyi* *moorloo* *wirrigawoo* *ngang-ji-mili*
fly eye bung:eye give-IT-CHAR
'The fly is a bung-eye giver.'

(9) *ward-ga-ngga* *thirri* *roorrij-birr+arni*
go-MD-ERG fight argue-3PL.NOM+ARNI
'They argued going along.'

Unfortunately, there is a serious problem here: it is evident that what specifies the Event in a finite VP is the collocation of a verbal root or stem together with a classifier, a bound morpheme that assigns the verbal lexeme into one of a set of about a dozen categories (see further McGregor 1990: 557, 2002: 58). It is not the root or stem by itself that fulfils the Event role. Thus in examples (6) and (7) it is the collocation of *ward-* 'go' and the classifier +I 'be, go' and *jard-* 'put flame to' and the classifier +DI 'catch', respectively, that serves to characterise the Event, not *ward-* 'go' and *jard-* 'put flame to' separately. In combination with different classifiers these two roots denote

7 Confusingly, FGG uses the same label (*Process*) for both the referent of a VP and the major phrasal role constituting the VP. In this paper I reserve the term *Process* for a VP referent, and distinguish this terminologically from the Event role.

different Event types: *ward-* ‘go’ in collocation with +A ‘extend’ denotes an event of carrying, while *jard-* ‘put flame to’ in collocation with the same classifier conveys the meaning ‘cook’ (roughly, ‘keep flame against’). It is the two items together, the verbal root or stem and the classifier, which jointly specify the designated event type. (See McGregor 2002 for more information on verb classification in Gooniyandi and other northern Australian languages.)

On the other hand, in non-finite clauses such as (8) and (9) the Event role in the non-finite VP is obviously realised by a verbal root or stem alone; the classifier morpheme does not appear. The same criterion, that is, categorises two different types of linguistic expression depending on clausal context, whether the clause is finite or non-finite. Only one of these expression types (the lexical verb root or stem) is the sort of unit we expect to be categorised into a parts-of-speech category, namely the expression type that is most restricted in distribution, the one that occurs in non-finite contexts.

Nominals, according to the FGG specification, can serve a role in an NP or in the Event role in a VP.[8] Which NP role is not specified: nominals differ in terms of the NP roles that they tend to serve, though there appear to be no absolute restrictions (except perhaps in the case of closed class nominals like determiners and numerals). Thus, *yoowooloo* ‘man’ in examples (1), (3), and (4) above serves in the Entity role; however it can equally serve in the Qualifier role, as in (10), even occasionally in the Classifier role, as in (11), where Banjo is categorised as an Aboriginal man.

(10) *ngidi-yoorroo* *yoowooloo*
1PL.EXC.NOM-DU man
‘we men’

(11) *yoowarni* *yoowooloo* *banjo /* *bililoona-ya /*
one man Banjo Bililuna-LOC
‘There was an Aboriginal man Banjo, at Bililuna.’

Yoowooloo ‘man’ can also appear in a VP, where it occurs in the same position as a verb root or stem, as shown by (12):

(12) *yoowooloo-ø+windi*
man-3SG.NOM+BINDI
‘He became a man.’

The third part-of-speech, adverbials, are even less restricted than nominals in terms of the grammatical relations they may realise, and include not just roles in NPs and VPs (like nominals), but also roles in clauses and postpositional phrases (PPs). The next two examples illustrate adverbials in clausal and PP usage, respectively: in (13) the adverbial *balngarna* ‘outside’ serves in a dependency relation of spatial enhancement (see next section, §2.2) to the core of the clause, indicating the location of the situation; in (14) *biliga* ‘middle’ and *ngiwayi* ‘south’ occur in comitative PPs, which in turn serve in relations of spatial enhancement to the clausal core.

---

8 Of course, the above qualifications concerning the fillers of this role must be taken into account regarding this statement.

(13) *niyaji-ya warang-birr+i / balngarna /*
this-LOC sit-3PL.NOM+I outside
'At that time/place they were sitting out in the open.'

(14) *niyaji-nhingi / thirroo-binyi baj-gi-wirr+i-yi biliga-ngarri /*
this-ABL kangaroo-PER get:up:and:go-INC-3PL.NOM+I-DU middle-COM
baabirri jirrinyoowa / ngiwayi-ngarri gamboornjoowa /
inside Jirrinyoowa south-COM Gamboornjoowa
'Then they went for kangaroos on the way; below Jirrinyoowa, to the south, at Gamboornjoowa.'

The unrestricted specification of adverbials given in FGG results in the inclusion of disparate groups of adverbials in the category, including manner adverbs, spatial adverbs and temporal adverbs. The adverbial category is thus a rag-bag of disparate types of item, not a unified lexical category. As will be shown in §2.2, the specification can be tightened up so as to distinguish the three major types of adverbials, which consequently emerge as more natural categories.

The FGG characterisation of the three main parts-of-speech includes no stipulation "without further morpho-syntactic measures" on the filling of roles by lexemes of the various parts-of-speech, as is the case in the FG and FDG theories of parts-of-speech. There is a reason for this. The "measures" referred to are processes giving rise to new lexical items, prototypically class-changing derivational morphology. Granted that the effect of such processes is the creation of new lexemes that can serve in particular grammatical relations, it is clear that the original (underived or measureless) root lexeme cannot be serving in that — or indeed any — grammatical relation in the derived form.

It might however be thought that such a stipulation would permit us to exclude the VP environment from the possible range of usages of nominals. Thus, it could be proposed that in examples such as (12) the classifier +BINDI is a "morpho-syntactic measure", and thus nominal stems may not themselves serve in the Event role in a VP. In this case nominals would be less flexible than presented above. The difficulty with this proposal is that it applies equally to verbals (recall the discussion following examples (8) and (9) above), yielding the implausible situation that verbals can only serve in the Event role in the most marked and infrequent grammatical context, non-finite clauses and VPs, and that "measures" need to be taken for them to be deployed in their unmarked environment of occurrence, finite clauses and VPs. I conclude that it is not viable to regard the attachment of a classifier as a "morpho-syntactic measure". A resolution of these difficulties is proposed in (§2.2), where the characterisation of verbals is refined.

To wind up this section, it is observed that according to the FGG parts-of-speech scheme, verbs are the most restricted part-of-speech in terms of their potential to realise grammatical relations (they correspond to a unique grammatical relation which they may realise in well-defined and restricted contexts); nominals are slightly less restricted, and adverbials even less restricted. This observation lends itself to interpretation in terms of markedness relations, more marked being associated with more restricted in distribution. This gives the scheme shown in (15), where > indicates 'is more marked than'; it can also be read as 'is more inflexible than'.

(15) verbals > nominals > adverbials

2.2 Revision and extension: Gooniyandi parts-of-speech in an SG framework

In this section I suggest a reformulation of the specifications for the parts-of-speech in Gooniyandi, one that embraces the other parts-of-speech categories identified in FGG (McGregor 1990: 138), and that avoids unmotivated rag-bag groupings such as adverbials and the unenlightening strategy whereby categories other than the three major ones are specified by listing their members, granted that they are effectively closed. This revision is situated within the parameters of SG (McGregor 1997). Before getting down to a detailed discussion of the parts-of-speech, it is essential that some basic features of this theory be presented. This is done in §2.2.1; following this, §2.2.2 presents the revised characterisation of the Gooniyandi parts-of-speech.

2.2.1 Brief overview of SG

Fundamental to SG is the notion that grammar is a semiotic system (McGregor 1997;[9] see also e.g. Goldberg 1995; Langacker 1987, 1991; Shaumyan 1987 for similar views). What this means is that grammatical entities of two primary types, units (also known as constructions, as per construction grammar — e.g. Goldberg 1995) and relations, are signs in the Saussurean sense: they consist of two inseparable and mutually defining components, a form (signifier) and a (coded) meaning (signified).

Units are of course language-specific, and are constituted by component units serving grammatical relations in or with one another. SG presumes that the units of any language can be classified into four generic types that can be placed on a hierarchy in terms of decreasing size: clauses, phrases, words, and morphemes.

According to the SG model four distinct and independent super-types of syntagmatic relations exist in the grammars of all languages: constituency, dependency, conjugational, and linking. Table 1 provides basic formal characterisations of these four super-types.

9 This does not, it should be stressed, mean that all aspects of the grammar of a language are semiotically significant. For instance, grammatical patterns that are invariant — for instance, the placement of prepositions in initial position in the NP (if this is the only option in a given language) — cannot possibly be semiotically significant since they cannot contrast with anything.

Formal type	Characterisation
constituency	part-whole relations; 'dominates', 'is the mother of'
dependency	part-part relationships, 'is a sister of'
conjugational	whole-whole relationships, 'encompasses'; one whole is a modification of the other, which is "nested" within it
linking	free relationships spanning units of arbitrary types and sizes

Table 1: *SG typology of grammatical relations*

Being grammatical signs, meanings are inherently associated with each of these relation types. These are summarised below; see McGregor (1997) for fuller elaboration and discussion.

Constituency relations code experiential meanings, and constitute the resources of a language for constructing representations of the world of experience — things, events, happenings, and so on. The major constituency relations in the Gooniyandi clause (and perhaps universally) are: State-of-Affairs (realised by a VP), Actor (realised by an NP that is cross-referenced by a NOM pronominal in the verb), and Undergoer (realised by an NP that is cross-referenced by an accusative pronominal in the verb).[10]

Dependency relations are part-part relations that code logical meanings, which construe the ways in which things, events, and happenings are related to one another; they provide the grammatical means of constructing a "logic" of interrelations among phenomena. Dependency relations are categorised according to two dimensions: (i) parataxis (equality) vs. hypotaxis (head and dependent); and (ii) extension (adding something), elaboration (restatement, reformulation, or addition of clarifying information), and enhancement (provision of circumstantial information).

Conjugational relations are whole-whole relations; here one whole is effectively included within the confines of another, resulting in a modification of it, shaping it in a way appropriate to its usage in the interaction. Conjugational relations construe interpersonal meaning, which is concerned with the construction and maintenance of social relationships amongst people; in this respect, language is construed as a mode of action, not a tool of thought. Again there are two primary independent dimensions in the classification of conjugational relations: (i) scope (in which one whole holds the other within its domain) vs. framing (where one whole encompasses the other, indicating it is to be taken as a demonstration rather than a description — as per Clark & Gerrig 1990); and (ii) illocutionary (concerning illocutionary force), attitudinal (where the speaker's attitude to something is expressed), and rhetorical (in which the information expressed

10 This is not a claim that the universal grammatical relations are identical cross-linguistically (see also Dryer (1997: 115-143). Clearly, as signs they cannot be anything but language specific: their form and meaning will differ from language to language. The claim here is merely that certain grammatical relations are sufficiently similar to one another to be regarded as in some sense the same, and identifiable across languages.

by the unit is integrated within the framework of information shared among the speech interactants).

Linking relations code meanings of the textural type, the functional/semiotic domain of language that concerns the structuring of language in use, that is, the structuring of discourse in relation to context, linguistic and extralinguistic. Linking relations provide structure to tokens of language and grammar, so that they "hang together" as coherent wholes. Textural relations include: (i) indexical relations, in which a linguistic item points to something, often outside of language; (ii) connective relations, in which two elements are bound together; (iii) marking relations, whereby grammatical relations or categories are labelled; (iv) covariate relations, relations that obtain between items in discourse by virtue of belonging to same semantic system; and (v) collocational relations, which obtain among items of text by virtue of statistical associations with one another, not by virtue of semantic commonality.

SG assigns a 4-dimensional "geometry" to grammar, corresponding to the four distinct types of grammatical relations. These four dimensions are effectively independent of one another, and the representation of the structure of a given clause or sentence in terms of a single dimension is thus typically descriptively inadequate.

Nonetheless, as it turns out, it is possible to typologise the clauses of Gooniyandi into four principal types according to the nature of the clause's inherent grammatical relations. (Of course, matters can be complicated by having non-inherent grammatical relations in the clause that are of different fundamental types to the inherent ones. This does not affect the typology, however.) Table 2 shows this typology.

Clause type	Semiotic	Meaning/function	Formal properties
Situational	Experiential	Refer to situations	Have an inherent SoA, and one or more inherent participant roles
Relational	Logical	Establish connections between entities and/or qualities	Verbless, two inherent roles
Minor	Interpersonal	Minor types of speech function — never express a proposition	Typically realised by a single interjection
Presentative	Textural	Presentative: present or introduce some entity into the text	Verbless, single inherent role

Table 2: *Gooniyandi clause types according to the semiotic components*

With a few qualifications, the grammatical relations (phrasal and clausal) identified in §2.1 above for Gooniyandi are also viable in the SG analysis.

2.2.2 SG characterisation of the Gooniyandi parts-of-speech

The SG characterisation of the Gooniyandi parts-of-speech system is according to both the type of grammatical relation that the lexeme may serve in and the type of grammatical unit it may occur in. In brief:

> VERBS are lexemes that must serve, either alone or in collocation with a classifying morpheme, in the Event role in a VP;
> NOMINALS are lexemes that have the potential to serve in a grammatical relation of some sort in an NP;
> ADVERBS are lexemes that may serve as attributive dependents in VPs;
> ADVERBIALS are lexemes that may serve as enhancing dependents in clauses;
> PARTICLES are lexemes that may serve in scopal relations (McGregor 1994), providing interpersonal modification of other linguistic units;
> SOUND EFFECTS are lexemes that are normally found in demonstrational usage (in the sense of Clarkand Gerrig 1990; see also McGregor 1994), and may serve as full clauses by themselves; they may not serve grammatical relations in other linguistic units;
> INTERJECTIONS are lexemes that must serve as full clauses by themselves, and may not serve in grammatical relations in clauses or smaller linguistic units.

As this listing shows, the primary parts-of-speech — verbs, nominals, adverbs, adverbials, and particles — are defined in terms of the range of types of grammatical relations the lexemes may serve in units of specified types. The two more minor categories are defined simply according to the type of unit they may comprise.

Table 3 attempts to render the above characterisations in a more intuitively appreciable way. (Note that for convenience of representation not all of the relevant grammatical relations are indicated; for instance, the Deictic role in NPs, an indexical relation, is omitted.) A primary division has been made between those parts-of-speech that are defined in terms of both relations and units, and those defined in terms of units alone (final two columns): these are the parts-of-speech that typically serve as minor clauses. Dark shading in a cell indicates the key grammatical relations and/or units that words of the relevant part-of-speech may serve in or as; words of each part-of-speech always have the potential to serve in these ways.

Grammatical relation								
Constituency			Dependency		Conjugation			
Unit type								
VP		NP	VP		Clause		Full clause	
Event, non-finite VP	Event, finite VP	Entity	Attri-bution	Attri-bution	Enhance-ment	Scope	As demon-stration	Minor speech function
Verb	Verb … classi-fier collo-cation							
		Nominal						
		???	Adverb					
					Adverbial			
						Particle		
							Sound effect	Interjection

Table 3: *Scheme of Gooniyandi part-of-speech*

As we have already seen, words of a given part-of-speech often have the potential to serve in other grammatical relations or units than the defining ones; these are shown in the table by light shading. It will be observed that in all instances these additional relations and/or units lie to the left of the defining ones in Table 3, provided we keep within the two primary categories of the parts-of-speech. The Gooniyandi parts-of-speech can thus be aligned on a hierarchy in such a way that items to the right of a given part-of-speech may (but do not necessarily) have the potential to "function as" items of the given part-of-speech category; they do not have the potential to "function as" parts-of-speech to the right on the hierarchy. This provides us with a somewhat more elaborated markedness hierarchy than the one shown in (15).

There are a number of gaps in my current understanding of the Gooniyandi parts-of-speech system. For one thing, we cannot interpret Table 3 as an implicational scale: it does not follow, for instance, that if an item is attested in a particular relation it will also have the potential to serve in roles to the left in the table; nor can we presume that a continuous chunk of the hierarchy will be represented by a given part-of-speech. These observations may or may not be consequences of inadequacies in the corpora; I suspect not, however.[11]

Table 3 provides some notion of the extent to which the Gooniyandi parts-of-speech categories are flexible. However, only a selection of grammatical relations and units of

11 In a few instances no more than a single lexeme or two from a given part-of-speech is attested in a grammatical relation to the left of the defining ones (see below). These are considered exceptional features of particular lexemes, and as such are not indicated in the table.

the language are shown, namely those that are relevant to the characterisation of at least one part-of-speech; thus the full extent of flexibility of the various parts-of-speech is likely be considerably greater than indicated in the table. It is an open question whether the hierarchical associations between the parts-of-speech, grammatical relations and units of Table 3 remains viable when the full range of grammatical relations and units are taken into account. In the remainder of this subsection we discuss in more detail the general properties of each part-of-speech in turn.

In the SG scheme, verbs remain the most highly marked part-of-speech category, being restricted in occurrence to VPs, where they must serve in the Event role either by themselves in collocation with a classifying morpheme. The former circumstance emerges, as we have seen, in the environment of non-finite clauses, where verbs — as shown by examples (8) and (9) above — serve in the Event role in the non-finite VP. The latter circumstance obtains in finite clauses, as shown by examples (6) and (7).

It must be admitted that verbs sometimes occur in environments that are not unequivocally either of the two indicated in Table 3. Thus in examples like (16), (17), and (18) the verb occurs in an environment that could be analysed as an NP: in (16) *dal-mawoo* 'extended' could perhaps be attributed of *barndi* 'arm'; in (17) *ward-nhingi* 'from walking' is plausibly attributed of *thinga* 'foot'; and in (18) *bayal-woo* 'for swimming' might be analysed as a PP in a dependency relation to the NP *gamba* 'water'. However, according to this analysis (16) would not involve the verb *dal-* 'extend' serving in a grammatical relation in the NP; rather the derived form *dal-mawoo* 'extended' would.[12] Thus this example is not problematic for our characterisation of verbs, regardless of whether non-finite clause or derivational morpheme analysis turns out to be preferable. No such explanation is readily available for (17) and (18) — the encliticised postpositions clearly do not derive new lexical items, and we are forced to adopt a non-finite clause analysis to preserve the above characterisation of verbs as a part-of-speech.

(16) [*barndi dal-mawoo*] *wilajka jarrk-ban-j+i*
arm extend-INF around jump-CNT-3SG.NOM+I
'(The brolga) was dancing around with its wings out.'

(17) [*thinga ward-nhingi*] *bagi-ri-ngarra* *banda-ya*
foot go-ABL lie-PRS/3SG.NOM/I-1SG.OBL ground-LOC
'My footprint lies on the ground.

(18) *gamba* [*bayal-woo*] *yijgawoo*
water swim-DAT bad
'The water is no good for swimming.'

12 The case for the derivational morpheme *-mili* 'characterised by' is perhaps even clearer. This morpheme can be attached to a verb, deriving a nominal stem, as in e.g. *moorniny-mili* (fuck-CHAR) 'promiscuous, randy', which is typically found in an attributive relation to an Entity nominal. In such instances, there is little evidence in support of non-finite clause status. Nevertheless, it is the derived form that serves in the dependency relation, and the verbal root itself discharges no grammatical relation whatever in any containing unit.

There is some independent supporting evidence for the non-finite clause status of the PPs in (17) and (18). Although there are no exactly parallel examples in which there is additional material present that uncontentiously belongs to a clause including the non-finite verb, there are a small number of apparently similar examples in which additional material occurs that evidently belongs to a non-finite clause containing the verb root or stem, which has a postposition encliticised to it. Thus, in (19) there is a *-nhingi* ABL marked non-finite clause that links to the main finite clause by a dependency relation of enhancement, comparable to the enhancing relation involved in the NP in (17). And in (20) what is marked by *-woo* DAT is clearly a non-finite clause, which serves in a relation to the main verbal clause comparable to the dependency relation in (18), namely the enhancing relation of purpose.

(19) *gamba ngoorloog-nhingi yalij-li+mi*
water drink-ABL get:sick-1SG.NOM+MI
'I got sick from drinking grog.'

(20) *gamba-ya nyoombool-woo malab-mi-ng+arni*
water-LOC swim-DAT tie:up-IT-1SG.NOM+ARNI$_1$
'I did up my hair for a swim in the water.'

Given that non-finite clauses are involved in these examples, it seems not unreasonable to presume that the same holds for examples like (17) and (18). That is, the verbs marked by postpositions in such examples represent highly elliptical non-finite clauses, that serve in the Event role in the VP of these non-finite clauses.

Nominals, as per Table 3, typically serve in the experiential role of Entity in an NP, or in the dependency relation of attribution to an Entity nominal (which corresponds to the FGG role of Qualifier). Many nominals are attested in both relations. However, as remarked above, there are differences among nominals as regards the preferred NP relations they serve in, whether they prefer the experiential role Entity or the dependency role of attribution. There are also a considerable number of nominals that are restricted to just one of the two relations. A number of nominals specifying animal and plant species, for instance, are attested only in the Entity role. On the other hand, numerals are attested only in relations of attribution never in the Entity role, and there is no evidence that they can serve this role.[13, 14]

Nominals are also encountered (as we saw in the previous section — recall example (12) in collocation with a verbal classifier, which collocation serves in the Event role in a VP. As far as I am aware, realisation of the Event role is possible only in finite clauses. One other environment in which nominals are sometimes found is in minor clauses, in as it were interjective or vocative usage. Thus, for example, the nominal *yoowooloo* 'man' can also be used as an attention-getting interjection, *yoowooloo!*

13 Numerals also occur in the Quantifier role; however, it is uncertain what type of grammatical relation this is in the SG scheme.

14 One possibility would be to restrict nominals to those items that can occur in the Entity role, and to assign those that only occur in attribution to a small restricted part-of-speech. I have no strong objections to such a proposal.

'man!'. Again, only a relatively small subset of nominals are attested in this environment, though one expects that given an appropriate context most nominals (with the probable exception of numerals) could be used in this way.

According to the above definition, pronominal like *ngidi* 'we plural exclusive' in (10) above are also nominals, since they may fill the Entity role in an NP. However, there are reasons to distinguish them as a separate subclass within the macro category of nominal. In particular, they are restricted in that they may not serve (in collocation with a classifying morpheme) in the Event role in a VP, unlike nominals generally. They may, however, serve as attributive dependents in NPs, as in:[15]

(21) *riwi* *ngirrangi*
place 1PL.EXC.OBL
'our country'

Adverbs, as characterised above, are lexemes that may serve as attributive dependents in VPs. This definition (which is in essence identical with the FDG definition) picks out manner adverbs, of which Gooniyandi has a fairly restricted set, possibly only around a score or so, which indicate speed, force, and other qualities of the event. (22) is illustrative.

(22) *barnbarra* *ward-j+i*
slowly go-3SG.NOM+I
'He walked slowly.'

Adverbs are not restricted to this grammatical relation, and may in addition serve (in collocation, of course, with a classifier) in the Event role in a VP, as in (23). Some also have the potential to attribute on other adverbs, as shown by the derived adverb stem *mayaarrayaarra* 'hard, energetically' in (24).

(23) *waya* *thirrkirli-ø+windi*
wire straight-3SG.NOM+BINDI
'The wire straightened.'

(24) *maya-arra-yaarra* *galjini* *girra-girra-y+i*
hard-MNR-RDP fast run-RDP-3SG.NOM+I
'He ran very quickly.'

There is some evidence that adverbs may even have the potential of serving in the Entity role in NPs. For instance, there are a few instances in which *galjini* 'fast' appears to be used to denote a horse race. (The analysis of these examples is somewhat uncertain, and it is possible that the adverb instead belongs to an elliptical non-finite clause.) In example (25) the derived adverb *maya-arra-yarra* (hard-MNR-RDP) 'hard, energetically' — which involves partial reduplication of the derived adverb *maya-arra*

15 It should be noted that the oblique forms of the pronominals are inflected case forms, associated with particular grammatical environments of usage, which effectively select them. They are not derived forms of the roots.

'hard'[16] — appears to be used first in Entity function in an NP, meaning something like 'speedy ones' or 'energetic ones'. It is also used subsequently in the same example as a dependent in an NP (the head of which is a nominal derived from an adverb), and as a dependent in a relational clause.

(25) *ngarragi yawarda / maya-arra-yaarra-yoorroo garndiwirri /*
my horse hard-MNR-RDP-DU two
galjin-gali-yoorroo / maya-arra-yaarra-yoorroo galjin-gali /
fast-GD-DU hard-MNR-RDP-DU quick-GD
maya-arra-yaarra ngarragi ya ... yawarda /
fast-MNR-RDP my hor horse
'My horses were both fast ones. They were two good racers; my horses were fast.'

For want of a better term, *adverbial* is used to specify the other lexical types grouped under adverbials in FGG, but which show quite different grammatical behaviour to adverbs, namely spatial and temporal adverbials (McGregor 1990: 155-164). These items normally serve as enhancing dependents on the nuclear situations specified by clauses, as illustrated by (26) and (27) respectively. Temporal adverbials allow little in the way of morphological modification, except for reduplication, while spatial adverbials come in a range of types, two of which admit inflections.

(26) *riwi ngirnda boorroonggoo bagi-y+i-nhi /*
camp this north.ALL lie-3SG.NOM-I-3SG.OBL
'His camp was to the north of him.'

(27) *wooboo-woorra / garra-garrwaroo boojoo-ø+windi / warrgoom /*
cook-3PL.NOM+A RDP-afternoon finish-3SG.NOM+BINDI work
'They heated (the branding irons), and late in the afternoon finished work.'

Temporal adverbials very occasionally serve in the Entity role in an NP, as in (28), and occasionally in the Classifier role (McGregor 1990: 262); however, they appear not to be able to serve in any other NP role. They can also serve (naturally, in collocation with a classifier) in the Event role in a VP, as shown by (29); this usage is quite rare, and attested for only a few temporal adverbials — though there is no reason whatever to suspect that there are any principled restrictions on temporal adverbials occurring in this context.

(28) *yaanya-ya garrwaroo*
other-LOC afternoon
'the other afternoon'

16 The root *maya* 'hard' itself is a nominal.

(29) *barrangga* *barrangga-wa-wani /* *wila*
build:up:to:wet build:up:to:wet-PROG-3SG.NOM/ANI finish
gad-birr+ini /
leave-3SG.ACC/3PL.NOM+BINI
'As it gets hot, they leave off work.'

Spatial adverbials can serve as elaborating dependents in NPs (example (30), and as (with a classifier) Events in finite VPs (example (31) and (32), but not, it seems, in the Entity role in NPs.

(30) *rarrin-gi+ri* *mirra* *baboorroonggoo*
hang-PRS+3SG.NOM/I head down.ALL
'(Flying foxes) hang with heads down.'

(31) *mirri* *laandi-ya-ø+woondi* *miga-ya* *bij-b+arni*
sun up-SUB-FUT+3SG.NOM+BINDI thusly-LOC emerge-2SG.NOM+ARNI
'Come when the sun is high.'

(32) *niyi-yangga* *jibirri-w+ani,*
that-ABL downstream-3SG.NOM+ANI
'Then he turned down (into a hole).'

Particles, which number about a dozen, prototypically serve in conjugational relations (McGregor 1997: 64-70, 209-210). Conjugational relations, as indicated above, code meaning of the interpersonal type, and modify the way a linguistic unit is to be "taken" interpersonally. In Gooniyandi, particles serve in just one type of conjugational relation, namely scopal relations; in other words they hold propositions (usually) within their scope, indicating the speaker's line on them. In this circumstance the two relevant wholes are the unmodified clause itself, and the clause together with the particle that has scope over it. Thus in (33) and (34) the particles *mangarri* 'not' and *yiganyi* 'uncertain' specify the speaker's line on the proposition expressed by the clause, respectively that it is not the case (that there was a homestead at the specified location) and that it is possible, though uncertain (that the speaker will come some night).

(33) *mayaroo* *mangarri* *wara-y+i* *miga-ya /* *marlami /*
house not stand-3SG.NOM+I thus-LOC nothing
'There was no homestead there then, nothing.'

(34) *yiganyi* *maningga* *ward-ja-wi+ng+i*
uncertain night go-SUB-FUT+1SG.NOM+I
'Maybe I'll come one night.'

Particles can also be found in other grammatical environments, in units other than major clauses. Some are attested in minor clauses, expressing minor speech act types (see Table 2). Thus, for instance, the particle *mangarri* 'not' is not infrequently used in this way, in denials, as shown by example (35). It is also used in responses to information questions, as shown by (36), where it does not, of course, serve to deny the proposition presumed by the question — in this instance, that the addressee is doing

something. Although only a few particles are attested in the context of minor clauses, there seems to be no reason in principle why any particle cannot be used in this way — that is, effectively as interjections.

(35) A: *nginyji joorloo ngoorloog-ji+ø+mi*
2SG.CRD too drink-2SG.NOM+3SG.ACC+MI
B: *mangarri nganyi marlami ngoorroo-yarndi yaabja-ngga*
not 1SG.CRD not that-PL some-ERG
ngoorloog-birr+ø+a
drink-3PL.NOM+3SG.ACC+A
A: 'You drank with them?'
B: 'No, not me. That other lot drank.'

(36) A: *yiniga-a+nggirr+a-rri*
do:something-PRS+2PL.NOM+A-PL
B: *mangarri maa ngab-girr+a*
not meat eat-2PL.NOM+3SG.ACC+A
A: 'What are you doing?'
B: 'Nothing. We are eating meat.'

Some particles at least have the potential of serving functions in NPs. For instance, *mangarri* 'not' in (37) and (38) evidently serves in NP roles.[17] In (37), the particle manifestly fulfils an interpersonal role: it has scope over the following numeral, and the two words together as a unit presumably serve in either a Quantifier or Qualifier role in an elliptical NP, meaning 'not one plane'. In (38), by contrast, *mangarri* 'not' would seem to serve in the Entity role, i.e. an experiential role rather than an interpersonal role. Again there seems to be no reason in principle why any particle would be precluded from the first usage as an interpersonal modifier in an NP. However, for their non-interpersonal usages in NPs, one has less grounds for confidence in presuming regularity, and these usages may well be idiosyncratic.

(37) *garndiwangoorroo, bin.gidi-ngarri, garndi:::wangoorroo*
many feather-COM many
mangarri yoowarni,
not one
'There were lots and lots of planes.'

(38) *mangarri-ya warang-j+i-wirrangi / mila-win+ø+a /*
not-LOC sit-3SG.NOM+I-3PL.OBL see-3PL.ACC+3SG.NOM+A
'Having nothing, he sat near them watching them.'

17 Note that this implies that the above characterisation of nominals requires some further fairly obvious restrictions in order to exclude particles from the part-of-speech.

Other non-interpersonal usages of particular particles also appear to be quite idiosyncratic and unsystematic. Thus in (39) *yiganyi* 'uncertain' indicates an attribute of the event, how it was performed — uncertainly, and thus by implication, sneakingly.[18]

(39) *wab-jing+i* *yiganyi-ngga-nyali*
sniff-3SG.NOM+DI uncertain-ERG-REP
'He sniffed it sneakingly.'

The final two parts-of-speech categories, sound effects and interjections, are relatively minor categories in the sense that they are numerically small, and their members enter into a very restricted range of syntagmatic relations with other linguistic units. Because of this behaviour they are categorised as paralinguistic elements in McGregor (1990: 138). Lexemes from both parts-of-speech typically occur as the sole components of minor clauses, with no other units in them, and uttered on their own intonation contours. For instance, the sound effect representing someone falling from a great height, *nrrr*, typically occurs by itself, on its own intonation contour. Interjections as in *yoowayi* 'yes' and *ba* 'come on, let's go' likewise typically occur independently on their own intonation contour.

The difference between sound effects and interjections is a functional one, concerning the speech function discharged by the minor clause type they typically comprise. Sound effects involve, as indicated above, demonstration of their referents, prototypically noises (as per Clarkand Gerrig 1990; see also McGregor 1994, 1997), rather than their description. Interjections, by contrast, express minor interpersonal speech functions such as offering (e.g. *nya* 'here you are'), confirming (e.g. *yoowayi* 'yes'), attention attraction (e.g. *yay*, *ngay* 'hey!'), and so on.

In the relatively few instances in which a sound effect or interjection seems to enter into syntagmatic relations with other linguistic units, as in (40) and (41),[19] I would argue that they represent separate minor clauses in syntagm with major clauses. Thus, examples such as (40) and (41) are complex sentences.[20]

(40) *landiwali* *wood* *wood / doomboo* *jijag-j+i /*
from:above hoot hoot owl speak-3SG.NOM+I
'From above an owl hooted.'

---

18 It may as well be that *yiganyi* 'uncertain' in examples such as this is serving as a secondary predicate on the Agent. The problem with this suggestion is that there is no independent evidence that the particle can occur in an attributive role in either an NP (as in e.g. 'the uncertain person') or in a verbless clause (e.g. 'he is uncertain').

19 In examples such as (40), it should be noted, the sound effect is typically uttered in a distinctive speech register, in contrast to delocutives as illustrated in example (42).

20 It will be observed that in (40) the sound effect occurs within the major clause, the components of which are discontinuous. This does not argue against the complex sentence analysis: quotations are not infrequently interpolated within the boundaries of their framing clauses of speech.

(41) *yoowayi / ngirndaji yaanya thangarndi /*
yes this other word
'Yes, this is another story.'

Sound effects — though not interjections — are occasionally incorporated within major clause types. However, this happens only when they are used delocutively (Benveniste 1958/1971), as in (42).[21] The most appropriate way of analysing such uses of sound effects is to regard the item in delocutive use as a separate lexical item, distinct from the sound effect or interjection. In keeping with this, in such uses sound effects are phonetically normalised and uttered in normal voice registers.

(42) *ganbirra glig-glig-gi+r+i*
eagle click-RDP-PRS+3SG.NOM+I
'The eagle clicks.'

2.3 Concluding remarks

To wind up the discussion of the Gooniyandi parts-of-speech a few general observations are in order.

First, as the above discussion indicates, three of the four types of grammatical relation identified in SG (McGregor 1997) are relevant to the revised part-of-speech classification. The fourth type, textural, does not figure in the classification. There is a reason for this: items serving in textural relations are prototypically grammatical rather than lexical.[22] It makes sense to open up the part-of-speech system to include grammatical morphemes as well as lexical roots (as in the FGG system), and it is not particularly difficult to do this. For instance, postpositions could be defined as bound morphemes that serve in the textural function of marking the grammatical relation of an NP within the unit it belongs to. For reasons of space, however, I have not attempted this expansion here.

Second, as we have seen, both the nature of the grammatical relation a lexeme may serve in and the size and type of unit in which it occurs are relevant to its part-of-speech classification. The major parts-of-speech prototypically serve functions in phrasal units, whereas it is the more minor ones that serve in clausal roles (adverbials and particles) or as full clauses by themselves (sound effects and interjections). One might wish to apply Occam's razor and attempt to economise in the characterisations by selecting just one characteristic as defining, either relations or unit types. Both of these reductive strategies run into serious problems. Grammatical relations on their own are inadequate to the task of defining parts-of-speech; units can't simply be avoided by employing the

21 I suspect that some interjections also admit delocutive usage — this is attested in Nyulnyul, for instance — and that the absence of tokens is an artifact of the incompleteness of the corpus.

22 Potential exceptions include, among other things: (a) verbal classifiers, which are — like derivational morphemes — somewhat intermediate between lexical and grammatical, and indeed can be traced back diachronically to lexical verbs; (b) determiners, which typically serve in a textural role in NPs (labelled Deictic above), but can also serve as dependents within NPs.

specific grammatical relations like Entity or Event, which unavoidably invoke the classification of larger units. The alternative strategy of defining parts-of-speech in terms of the units a lexeme may occur in, ignoring grammatical relations, initially appears more promising, but in the end also runs into the same sort of difficulty. The reason is similar: verb phrases and noun phrases are distinguished by the different grammatical relations these two unit types serve in clauses: specifically, VPs are restricted to the SoA role, which role NPs may not serve in.

Third, the extent to which other grammatical relations than those discussed above may be served by items of the specified category is uncertain. That is, it is unclear (largely due to inadequacies in the data) whether it is just a small fraction of the items of the part-of-speech that may serve in other grammatical relations, or potentially the entire category. Of course, the greyed cells in Table 3 give some indication of the degree of flexibility of the Gooniyandi part-of-speech system; they give an indication of the extent to which multiple grammatical functions are associated with the distinct parts-of-speech categories. However, the full story of flexibility can only be appreciated when one takes into account the full set of grammatical relations distinguished in the language, and the full range of unit types, rather than the subsets defined by those items that are prototypically associated with a part-of-speech category. For instance, at least some spatial adverbials regularly enter into dependency relations with NPs, with which they form complex units, as shown by (43). These are dependency relations of enhancement (see further McGregor 1990: 287-289).

(43) *rirringgi* *gamba-ya* *ward-j+i*
side water-LOC go-3SG.NOM+I
'He went along the side of the water.'

These additional grammatical relations are clearly relevant to the full story of flexibility of the Gooniyandi parts-of-speech categories. It is, however, beyond the scope of this article (and the available information on Gooniyandi) to deal in depth with the full set of grammatical relations and unit types, and their relations with the parts-of-speech.

3. The vague semantics of lexemes in Gooniyandi

I have elsewhere (e.g. McGregor 1990, 1997) argued for a monosemic approach to lexical semantics in general, and in Gooniyandi in particular, whereby lexical items have quite abstract and non-specific semantic specifications (see also Ruhl 1989 and, with some qualifications, Levinson 2000). Lexical semantics thus differs in degree rather than kind from the semantics of grammatical items: the latter is typically more abstract and schematic than the former.

This view accords with the approach to parts-of-speech adopted in this paper: items that are flexible in the sense that they may discharge more than one grammatical function (without being derived) cannot have highly specific semantics. For instance, the semantics of a nominal such as *yoowooloo* 'man' cannot include entity specification, since the word can be used in a variety of grammatical relations which conflict with entity specification, as illustrated by examples (10), (11), and (12). It can also be used, as we have seen, in a minor clause, as an attention-getting interjection, *yoowooloo!* 'man!'.

I would argue that the coded meanings of phrases and clauses in Gooniyandi are in general derivable compositionally from the abstract coded meanings of the lexical and grammatical items making it up, along with the coded meanings of the grammatical relations and constructions.[23] Thus in (1), (3), and (4), it is the combination of the abstract semantic meaning of *yoowooloo* 'man' (which specifies nothing about entity status) plus the grammatical relation Entity (in the NP) that permits the interpretation that it is a male human being that is being referred to. In (10), by contrast, *yoowooloo* 'man' does not serve in the Entity role in an NP, but rather in an attributing dependency relation to the referential NP *ngidi-yoorroo* 'we two', this invoking the quality interpretation, that we are men.

Similar remarks apply for examples such as (34) and (39): the particle *yiganyi* 'uncertain' is specified semantically in a vague and abstract way, consistent with both interpersonal and dependency grammatical relations — that is, with either modification of the proposition expressed by a clause, or modification of the manner of performance of an event (or alternatively a mental attribute of the Actor, depending on the appropriate analysis of examples like (39). Similarly for other instances of flexibility.

Aside from the semantic compositionality illustrated in these examples, we also need to employ pragmatic implicatures in the interpretation of utterances, resulting in additional non-coded meanings. These inferred meanings perhaps include things like 'not all' if *some* is employed.

As argued in McGregor (1990, 2002), the coded meanings of Gooniyandi verbs are also radically abstract and vague, consistent with the fact that different categorisations of a single verb result in different coded meanings, and that verbs occurring without classifiers in non-finite VPs admit a variety of interpretations. These differences are in terms of Aktionsart, valency, and vectorial configuration (an abstract actional scheme for the event). This can be seen by comparison of the following examples, which show clearly that *mila-* 'see' cannot involve in its semantic specification valency or Aktionsart information:

(44) *nganyi* *mila-ngi+ri*
1SG.CRD see-1SG.NOM+PRS/I
'I am looking.'

(45) *nganyi-ngga* *mila-ng+arni*
1SG.CRD-ERG see-1SG.NOM+ARNI
'I saw myself.'

---

23 There are exceptions, places where semantic compositionality fails. This occurs in classification and stem derivation. As per McGregor (2002) in classification the semantics of a classifying lexeme is effectively irrelevant to the semantics of the classified lexical item. What is relevant is the semantics of the category marked by the lexeme; but even this meaning cannot necessarily be used compositionally to obtain the meaning of a classified item — there is a degree of arbitrariness in grammatical classification that prevents this (McGregor 2002). Similarly, as is well known, derivational processes are not entirely systematic semantically, and the semantics of a derived lexeme is only partly predictable from the meaning of its components.

(46) *nganyi-ngga wayandi mila-l+ø+a*
1SG.CRD-ERG fire see-1SG.NOM+3SG.ACC+A
'I saw a fire.'

(47) *nganyi-ngga mawoolyi-yoo mila-li+mi-wirrangi*
1SG.CRD-ERG children-DAT see-1SG.NOM+MI-3PL.OBL
'I glanced at the children.'

The conceptualisation of vague semantics of Gooniyandi lexical items is distinct from that of FDG as per e.g. Rijkhoff (2008), which builds on Wilkins (2000). In my view the coded semantics includes just those features that are recurrent in all uses of a unit (or relation), except in those highly exceptional cases where a pragmatic implicature might be invoked to negate it. Thus it is not a matter of highlighting relevant components of meaning and downplaying others, as per Rijkhoff (2008: 731). Rather, the components that can be downplayed are not there in the lexical meaning at all, and those that are highlighted correspond to meanings that are coded elsewhere, e.g. in the larger construction, or in the grammatical relations.

4. Conclusions

It seems to me that the FG and FDG approaches to parts-of-speech systems in the world's languages are essentially on the right track. In particular, the best way to characterise the parts-of-speech is in regard to the grammatical relations their members discharge and the nature of the grammatical unit they belong to or constitute; pure morphological criteria (such as the paradigms their members enter into, as per the received approach to parts-of-speech in Australian Aboriginal languages outlined in §1), and/or "semantics" (by which is meant intuitive conceptual senses, not genuine semantics in the sense of coded meanings) do not provide promising approaches. I have suggested that this general type of approach works well in Gooniyandi in regard not just to the three major parts-of-speech characterised in this way in FGG, but also in regard to the other categories of lexical items, which, due to their being relatively closed classes, were defined by listing in McGregor (1990). Moreover, this approach can be extended to encompass the grammatical or function units in the language, such as postpositions, verbal classifiers, tense, mood and aspect markers, and so on. I have not attempted this extension here, however, due to considerations of space. The main differences from the FG and FDG approaches concern matters of detail, specifically the grammatical relations and units distinguished.

The notion of flexibility provides a better story than a commonly invoked alternative, according to which zero-derivation is involved in cases such as the "verbal use" of nominals such as *yoowooloo* 'man' or the "modifier" use of particles like *yiganyi* 'uncertain'. The notion of zero derivation is highly problematic, and it is not at all clear that such zeros can be motivated by viable language-internal criteria (see e.g. Haas 1957; McGregor 2003). In particular, they run foul of the crucial requirement, which specifies that a genuine zero cannot be identified if as a result we are forced to the position that there is a contrast between zero and nothing.

To wind up the paper, I briefly comment on two further issues raised by the discussion of this paper. The first concerns correlations between parts-of-speech systems and other features of the grammar of particular languages: could it be that

apparent differences in the flexibility of parts-of-speech in north-western vs. eastern Australian languages (prototypical Pama-Nyungan languages) correlate with other typological differences? One possibly relevant typological difference concerns the inherent transitivity or valency of verbs. In Gooniyandi lexical verbs are radically ambicategorial, and are quite unspecified for valency and show few if any restrictions in terms of the transitivity of the clauses in which they may occur (see further McGregor 2002), whereas in the prototypical eastern Pama-Nyungan language verbs are allegedly inherently specified for valency, and are rigidly bifurcated into transitive and intransitive classes (Dixon 1980). If this is indeed a genuine typological difference, it would appear that the most inflexible part-of-speech, the verb, in Gooniyandi is relatively flexible vis-a-vis the most inflexible part-of-speech in eastern languages. This relative flexibility might reasonably be expected to extend to the less inflexible parts-of-speech in Gooniyandi, which accordingly might be expected to be more flexible than their correlates in eastern languages.

Second, the question arises as to what extent the SG approach adopted in this paper specifically for the description of Gooniyandi is generalisable to other languages, and can be deployed in a general typology of parts-of-speech systems. SG recognises a considerable number of specific grammatical relations (falling into the four super-types, and depending on the types of unit related), and it is an open question as to whether a particular subset of these relations can be selected in a principled way as inherently relevant to the definition of parts-of-speech systems, and a complementary subset excluded as irrelevant. Only if this is possible would it make sense to attempt to devise a hierarchy or implicational "scale" (perhaps two dimensional) of parts-of-speech systems along the lines of recent work in FDG, e.g. (van Lier 2008, 2009). The possibility that this would be viable seems slim to me.

References

Benveniste, E. (1958/1971). 'Delocutive verbs', in E. Benveniste (ed.), *Problems in general linguistics*. Coral Gables: University of Miami Press**:** 239-246.

Blake, B. J. (1987). *Australian Aboriginal grammar*. Sydney: Croom Helm.

Clark, H. H.and R. J. Gerrig (1990). 'Quotations as demonstrations', *Language* 66: 764-805.

Dixon, R. M. W. (1980). *The languages of Australia*. Cambridge: Cambridge University Press.

Dixon, R. M. W. (2002). *Australian languages: their nature and development*. Cambridge: Cambridge University Press.

Dryer, M. S. (1997). 'Are grammatical relations universal?', in J. Bybee, J. Haimanand S. A. Thompson (eds.), *Essays on language function and language type: dedicated to T. Givón*. Amsterdam and Philadelphia: John Benjamins, 115-143.

Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago and London: University of Chicago Press.

Haas, W. (1957). 'Zero in linguistic description', *Studies in linguistic analysis*. Special Publication of the Philological Society. Oxford: Blackwell, 33-53.

Hengeveld, K. (1990). 'A functional analysis of copula constructions in Mandarin Chinese', *Studies in Language* 14 (2): 291-323.

Hengeveld, K. (1992). 'Parts of speech', in M. Fortescue, P. Harderand L. Kristoffersen (eds.), *Layered structure and reference in a functional perspective: papers from the*

functional grammar conference in Copenhagen 1990. Amsterdam: John Benjamins, 29-55.
Hengeveld, K.and J. L. McKenzie (2008). *Functional Discourse Grammar: a typologically-based theory of language structure*. Oxford and New York, Oxford University Press.
Hengeveld, K., J. Rijkhoff,and A. Siewierska. (2004). 'Parts of speech systems and word order', *Journal of Linguistics* 40 (3): 527-570.
Langacker, R. W. (1987). *Foundations of cognitive grammar. Volume 1*. Stanford: Stanford University Press.
Langacker, R. W. (1991). *Foundations of cognitive grammar. Volume II: descriptive applications.* Stanford: Stanford University Press.
Levinson, S. C. (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge, Massachusetts and London: The MIT Press.
McGregor, W. B. (1990). *A functional grammar of Gooniyandi*. Amsterdam: John Benjamins.
McGregor, W. B. (1994). 'The grammar of reported speech and thought in Gooniyandi', *Australian Journal of Linguistics* 14 (1): 63-92.
McGregor, W. B. (1996a). 'Arguments for the category of verb phrase', *Functions of Language* 3 (1): 1-30.
McGregor, W. B. (1996b). 'Attribution and identification in Gooniyandi', in M. Berry, C. Butler, R. Fawcett and G. Huang (eds.), *Meaning and form: systemic functional interpretations. Meaning and choice in language: studies for Michael Halliday*. Norwood: Ablex, 395-430.
McGregor, W. B. (1996c). 'The pronominal system of Gooniyandi and Bunuba', in W. B. McGregor (ed.), *Studies in Kimberley languages in honour of Howard Coate*. München & Newcastle: Lincom Europa, 159-173.
McGregor, W. B. (1997). *Semiotic grammar*. Oxford: Clarendon Press.
McGregor, W. B. (2002). *Verb classification in Australian languages*. Berlin and New York: Mouton de Gruyter.
McGregor, W. B. (2003). 'The nothing that is, the zero that isn't', *Studia Linguistica* 57 (2): 75-119.
McGregor, W. B. (2006). 'Focal and optional ergative marking in Warrwa (Kimberley, Western Australia)', *Lingua* 116 (4): 393-423.
Rijkhoff, J. (2007). 'Word classes', *Language and Linguistics Compass* 1 (6): 709-726.
Rijkhoff, J. (2008). 'On flexible and rigid nouns', *Studies in Language* 32 (3): 727-752.
Ruhl, C. (1989). *On monosemy: a study in linguistic semantics*. Albany: State University of New York Press.
Shaumyan, S. (1987). *A semiotic theory of language*. Bloomington and Indianapolis: Indiana University Press.
van Lier, E. (2008). 'Parts-of-speech systems and dependent clauses: a typological study', *Folia Linguistica* 40 (3-4): 239-304.
van Lier, E. (2009). *Parts of speech and dependent clauses: a typological study*. PhD thesis, University of Amsterdam.
Wilkins, D. P. (2000). 'Ants, ancestors and medicine: a semantic and pragmatic account of classifier constructions in Arrernte (Central Australia)', in G. Senft (ed.), *Systems of nominal classification*. Cambridge: Cambridge University Press, 147-216.

Jack-of-all-trades: the Sri Lanka Malay flexible adjective

Sebastian Nordhoff

1 Introduction

Discussions of languages with very flexible parts-of-speech systems have focused on monocategorial languages in the past, i.e. languages where all lexemes are said to be in the same word class. Examples are Samoan (Mosel and Hovdhaugen 1992), Riau Indonesian (Gil 1994), Tongan (Broschart 1997), Mundari (Croft 2005, Hengeveld and Rijkhoff 2005), Kharia (Peterson 2005), and Late Archaic Chinese (Bisang 2008a, 2008b). According to these authors, it is impossible to establish morphosyntactic differences between the lexemes in these languages. They are all in the same form class, and no other lexemes are in another class.

This focus on monocategorial languages has obscured the possibility that maximally flexible parts of speech can also exist alongside other lexical categories more limited in their potential.[1] For instance, it would be imaginable that a language has two word classes: a rigid category of verbs and a class of maximally flexible lexemes (i.e. lexemes that could immediately be used in verbal, nominal, adjectival and adverbial functions). Such a language would then not be monocategorial, but polycategorial. This paper presents the polycategorial system of Sri Lanka Malay, where a maximally flexible class of adjectives coexists with the more restricted categories of nouns and verbs.

In Sri Lanka Malay, distributional properties clearly indicate the existence of three morphosyntactically different form classes: verbs, nouns and adjectives. Verbs and nouns are limited in their potential use for different discourse functions. Such a limitation does not hold for adjectives. The clear distributional distinction between the three word classes sets Sri Lanka Malay apart from other Malay (and Austronesian) varieties. The reduction of flexibility of the noun class and the verb class is analyzed here as a result of extended contact with the adstrate languages Sinhala and Tamil, which both have a rigid parts-of-speech system.

The paper is structured as follows: After a brief introduction to Sri Lanka Malay (Section 2), we provide an overview of its word classes and their morphosyntactic characteristics (Section 3). Semantic aspects of the various word classes are discussed in section 4. Section 5 is concerned with the discourse functions of members of the various classes in Sri Lanka Malay and section 6, finally, deals with issues such as language contact and the diachronic development of the SLM parts-of-speech system.

2 Sri Lanka Malay

Sri Lanka Malay (SLM) is the language of the ethnic group of the Malays in Sri Lanka. They were brought by the colonial powers of the British and the Dutch between the 17th and the 19th century. While the lexicon has remained quite stable over time, the grammar has been heavily influenced by the adstrates Sinhala (Indo-Aryan) and Tamil (Dravidian) (Adelaar 1991, Smith et al. 2004, Bakker 2006, Slomanson 2006, Ansaldo 2008, Nordhoff 2009).

The immigrants were mainly soldiers, some exiled princes, and very few convicts and slaves. The mercenaries hailed from all over what is today Indonesia and Malaysia, but were recruited in Jakarta and later Penang (Hussainmiya 1990). In the regiment, some variety of Trade Malay local to Jakarta was therefore most probably the dominant linguistic code (Vlekke 1943, Paauw 2004, Smith et al. 2004, Smith and Paauw 2007).

[1] The existence of lexical categories with limited flexibility alongside more rigid classes has received more attention, e.g. Bhat (1994).

While the community was quite close-knit until the middle of the 19th century, the dissolution of the Malay regiment in 1873 led to a dispersion of the community. This and the Sinhala nationalist policy following the independence of Ceylon from the UK in 1948 led to attrition and language loss, which we are witnessing today.[2]

Today, the Malays make up 0.3% of the Sri Lankan population. They live in several towns in the Central and Southern Sinhala speaking area, were they make up between 1 and 3% of the local population (30% in Hambantota, over 90% in the Southern hamlet of Kirinda (Bichsel-Stettler 1989, Ansaldo 2005).

All Sri Lankan Malays can understand each other without problems, but there is a lot of idiolectal variation, which does not pattern geographically, but rather according to family lines (which may be present in several places).

3 Structuralist approach

In the approach established by the American structuralist in the beginning of the last century (e.g. Bloomfield 1933), form classes are defined by the distributional properties of lexemes (both morphological and syntactic). Lexemes with the same morphosyntactic properties are grouped in the same class. I will discuss the defining properties of SLM verbs, nouns and adjectives in turn. This classical take will be contrasted with a functional discourse-pragmatic analysis in the next section.

3.1 Verbs

SLM lexemes of the Verb class share the ability to take verbal prefixes, verbal preclitics, the nominalizer *-an* and a certain negation pattern. They are distinguished from nouns by the inability to co-occur with the indefiniteness marker *atthu*, the deictics *ini* and *itthu* and the plural marker *pada*. I will give an extensive overview of the properties of verbs, which will be important for the definition of adjectives later on.

3.1.1 Prefixes

The following prefixes can be combined with verbs in SLM:
- the past markers *su-*, examples (1) and (2) and *anà-*, examples (3) and (4)
- the conjunctive participle marker *asà-*, examples (5) and (6)
- the infinitive marker *mà-*, examples (7) and (8)
- the past tense negator *thàrà-*, examples (9) and (10)
- the non-past marker *arà-*, examples (11) and (12)
- the irrealis marker *anthi-*, examples (13) and (14).

The following examples show the use of these prefixes on verbs in constructed sentences (first) and natural discourse from the corpus (second), available at the Dobes-archive.[3] [4]

---

[2] Documentation of the language is currently being carried out as part of a Dobes-project consisting of Umberto Ansaldo, Lisa Lim, Walter Bisang and Sebastian Nordhoff.

[3] The examples are in a practical orthography. <c> and <j> represent palatal stops, <t> and <d> represent retroflex stops, <th> and <dh> represent dental stops. See Nordhoff (2009) for more information on SLM phonology.

[4] I would like to thank Tony Salim and Izvan Salim for going through all the examples again and suggesting improvements.

Sentences starting with eli... are taken from the corpus, but are elicited. These are taken if no (appropriate) example could be found in the corpus.[5]

(1) *Incayang* ***su-maakang***
he PAST-go
'He ate.' (K081112eli01)

(2) *Kitham=pe aanak pada=le karang bae=nang cinggala* ***su-blaajar***
1PL=POSS child PL=ASSOC now good=DAT Sinhala PAST-learn
'Our children have learnt Sinhala well now.' (K051222nar05)

(3) *Incayang* ***anà-maakang***
he PAST-go
'He ate.' (K081112eli01)

(4) *Seelon=nang lai hathu kavanan* ***anà-dhaathang***
Ceylon= DAT other INDEF group PAST-come
'Another group came also to Ceylon.' (K060108nar02)

(5) *Incayang* ***asà-maakang*** *su-miinung*
he CP-eat PAST-drink
'He ate and drank.' (K081112eli01)

(6) *Nyaakith oorang pada* ***asà-pii*** *thaangang arà-cuuci*
sick man PL CP-go hand NONPAST-wash
'The patients go and wash their hands.' (K060116nar03)

(7) *Incayang=nang* ***mà-maakang****(=nang) thàràboole*
he=DAT INF-eat=DAT cannot
'He cannot eat.' (K081112eli01)

(8) *Derang pada=nang atthu=le* ***mà-kijja****=nang thàràboole*
3PL PL=DAT one=ASSOC INF-make= DAT cannot=because
'They could not do a single thing.' (N060113nar01)

(9) *Incayang* ***thàrà-maakang***
he NEG.NONPAST-eat
'He did not eat.' (K081112eli01)

(10) *Puaasa muusing* ***thàrà-duuduk****=si?*
fasting season NEG.PAST-stay=INTERR
'You were not here during the fasting period, were you?.' (B060115cvs03)

(11) *Incayang* ***arà-maakang***
he NONPAST-eat
'He is eating.'

[5] The source for every example is given after the translation. Most text from which examples are drawn are available in the dobes-archive at http://corpus1.mpi.nl/ds/imdi_browser/.

(12) *Uncle, mana=ka **arà-duuduk**?*
uncle where=LOC NONPAST-live
'Where do you live?' (B060115cvs16)

(13) *Incayang **anthi-maakang***
he IRR-eat
'He will eat.'

(14) *Kithan=le **anthi-banthu***
1PL=ASSOC IRR-help
'We will help, too.' (B060115nar02)

3.1.2 Negation

As we will see later on (Section 3.3.6), negation patterns provide good cues to the class-membership of lexemes. SLM verbs show the following negation pattern: in the past tense, they are negated by *thàrà-* in examples (9) and (10) above. In the non-past tenses, they are negated by *thama-* (15) and (16) below.

(15) *Incayang **thama-maakang***
he NEG.NONPAST-eat
'He does/will/would not eat.'

(16) *Kitham=pe aanak pada **thama-oomong***
1PL=POSS child PL NEG.NONPAST-speak
'Our children don't speak (Malay).' (G051222nar01)

3.1.3 Nominalizer -an

The derivational suffix *-an* can be used to derive nominals from verbs, like *maakang* 'eat' → *makanan* 'food', *thaksir* 'think' → *thaksiran* 'thought'.

3.1.4 Negative features

So far, we have dealt with morphemes whose presence indicates that the lexeme they combine with belongs to the class of verbs. We now turn to morphemes whose co-occurrence with a lexeme precludes membership of that lexeme in the class of verbs. There are four main morphemes which can be used to exclude a lexeme from the verb class: the indefiniteness markers (*h*)*a(t)thu* , the deictics *in(n)i* or *it(t)hu* or the plural marker *pada*. The examples in (17) show the ungrammaticality of these combinations.

(17) a. **at(t)hu maakang*
INDEF eat
b. **i(n)ni/*it(t)hu maakang*
PROX/DIST eat
c. **maakang pada*
eat PL

To make the sentences in (17) grammatical, the nominalizer *–an* has to be added.

3.2. Nouns

Lexemes of the class of nouns are the mirror image of verbs. They can combine with the morphemes which are ruled out for verbs just discussed (Section 3.1.4), and they cannot combine with the verbal prefixes or verbal preclitics discussed in Section 3.1.1. The morphemes that indicate the membership of a lexeme in the class of nominals are:

- indefinite article *(h)a(t)thu*
- the deictics *in(n)i* and *i(t)thu*
- the plural marker *pada*
- the nominal negator *bukang*.

3.2.1. Indefiniteness marker

The indefiniteness marker can combine with nouns, where it can occur preposed (18-19), postposed (20-21) or in both positions (22).

(18) *Ini* ***atthu*** ***ruuma***
this INDEF house
'This is a house.' (K081112eli01)

(19) *Sindbad the Sailor* ***hatthu*** ***Muslim***
Sindbad the sailor INDEF muslim
'Sindbad the Sailor was a Moor.' (K060103nar01)

(20) ***Ruuma atthu*** *su-aada*
house INDEF PAST-exist
'There was a house.' (K081112eli01)

(21) *See* ***awuliya atthu*** *su-jaadi*
1s saint INDEF PAST-become
'I have become a saint.' (K051220nar01)

(22) *Sithu=ka* ***hathu*** ***maccan hathu*** *duuduk aada*
there= LOC INDEF tiger INDEF stay EXIST
'A tiger stayed there.' (B060115nar05)

3.2.2 Deictics

Both the proximal deictic *in(n)i* in examples (23) and (24) and the distal deictic *it(t)hu* in examples (25) and (26) can combine with lexemes of the noun class. They are always preposed.

(23) ***Ini ruuma*** *bìssar*
this house big
'This house is big.' (K081112eli01)

(24) ***Inni oorang****=nang* *itthu* *thàràthaau*
PROX man= DAT DIST NEG.know
'The man didn't know that.' (K07000wrt01)

(25) ***Itthu ruuma*** *pada* *bìssar*
those house PL big
'Those houses are big.' (K081112eli01)

(26) ***Itthu watthu****=ka* ***itthu nigiri*** *pada=ka* *arà-duuduk*
DIST time= LOC DIST land PL=LOC NONPAST-stay
'At that time, (they) lived in those countries.' (N060113nar01)

3.2.3 Plural marker

The plural marker *pada* can occur with many nouns (25) and (26), but not all of them, due to the well-known distinction between count nouns and mass nouns. Co-occurrence with a lexeme is then positive evidence for membership of that lexeme in the class of nouns, but impossibility to co-occur does not provide evidence to the contrary.

3.2.4 Negation pattern

Nouns are always negated by *bukang* when used as predicates.[6] The reverse it not true. While nouns are always negated by *bukang*, negation by *bukang* does not entail that we must be dealing with a noun. *Bukang* can also be used for constituent negation and clausal negation (cf. Gil this volume for the use of *bukan(g)* in other varieties of Malay). Therefore, this feature is only a cue to class-membership, but not sufficient in itself.

(27) *Incayang* *(hatthu)* *doktor* *bukang*
3S INDEF doctor NEG.NONV
'He is/was not a doctor.' (K081105eli02)

3.2.5 Negative features

Nouns are distinguished from verbs by their inability to combine with the verbal prefixes, the nominalizer *-an* or verbal negation patterns listed above. Example (28) shows the ungrammaticality of these combinations.

[6] Existential negation is done by *thraa*, e.g. *Bìrras thraa* 'There is no rice'.

(28) a. **su/*anà-/*mà-/*arà-/*anthi-* *ruuma*
PAST-/PAST-/INF-/NONPAST-/IRR- house
b. **ruma -(h)an*[7]
house-NMLZR
c. **thàrà-/*thama-* *ruuma*
NEG.PAST/NEG.NONPAST house

3.3 Adjectives

We now turn to the category crucial to this paper, the adjective. In order to signal its special status, and to make it clear that there are significant differences with adjectives in other languages, I will use the label 'Flexible Adjective' for this class in the following. This term should not be interpreted as entailing the existence of non-flexible adjectives in SLM. All SLM adjectives are flexible. Lexemes of this class can combine with all features characteristic of the Verb class and with all features characteristic of the Noun class. There is not one morphosyntactic process Flexible Adjectives could not take part in. Adjectives are the Jack-of-all-trades of SLM parts of speech. The following examples show that lexemes of the Flexible Adjective class can take all the morphology typical of the classes of verbs and nouns. The most prominent feature of Flexible Adjectives is thus their versatility.

3.3.1 Verbal prefixes

I have discussed above that a number of prefixes provide good cues to the membership of a lexeme in the class of verbs. This has to be amended, because Flexible Adjectives can combine with the very same prefixes. This means that these prefixes do not establish that a lexeme is a verb, because it could also be a Flexible Adjective. What these prefixes do establish is that the lexemes under discussion cannot be a noun. We list the verbal prefixes again for convenience and give examples of their use with Flexible Adjectival lexemes.

- the past marker *su-*, examples (29) and (30), and *anà-*, examples (31) and (32);
- the conjuctive participle marker *asà-*, examples (33) and (34);
- the infinitive marker *mà-*, examples (35);
- the non-past marker *arà-*, examples (36) and (37);
- the irrealis marker *anthi-*, examples (38) and (39);
- the negators will be discussed in a special section, 3.3.6.

(29) *Oorang* ***su-thiiňggi***
man PAST-tall
'The man grew up.' (K081112eli01)

[7] The nominalizer *–an* is realized as *–han* in certain contexts. Both forms are ungrammatical here.

(30) *Hatthu spuulu liimablas thaavon=na jaalang blaakang,*
INDEF ten fifteen year= DAT go after

inni kumpulan ***sa-mampus***
PROX association PAST-dead
'About ten, fifteen years after that, the association became defunct.'

(31) *Oorang* ***anà-thiiňǧgi***
man PAST-tall
'The man grew up.' (K081112eli01)

(32) *Itthu=nam blaakang=jo, kitham pada* ***anà-bìssar***
DIST after= EMPH 1PL PL PAST-big
'After that, we grew up.' (K060108nar02)

(33) *Aanak asà-bìssar* ***su-pii***
man CP-big PAST-go
'After the child had grown up he went away.' (K081112eli01)

(34) *Aanak pada asà-bìssar, skul=nang* ***anà-pii***
child PL CP-big school= DAT PAST-go
'After the children had grown up, they went to school.' (K051222nar04)

(35) *Oorang mà-bìssar=nang* ***arà-maakang***
man INF-big= DAT NONPAST-eat
'The man eats to become big.' (K081112eli01)

(36) *Oorang* ***arà-thiiňǧgi***
man NONPAST-big
'The man is growing.'

(37) *Ruuma* ***arà-kiccil***
house NONPAST-small
'The houses are getting small.' (K051222nar04)

(38) *Oorang* ***anthi-thiiňǧgi***
man RR-big
'The man will grow.'

(39) *Ithukapang gaathal* ***anthi-kuurang***
then itching IRR-little
'Then the itching will become less.' (K060103cvs02)

3.3.2 Nominalizer -an

Flexible Adjectives can be nominalized by *-an*, like *maanis* 'sweet' →*manisan* 'sweets', *sìggar* 'healthy' → *sìgaran* 'health'.

3.3.3 Nominal indefiniteness

The preceding sections have shown that Flexible Adjectives can combine with all verbal morphology.[9] We now examine to what extent Flexible Adjectives can combine with nominal morphology. One characteristic of nouns mentioned was the possibility to co-occur with the indefiniteness marker *atthu*. Example (40) shows that this marker can also co-occur with Flexible Adjectives.

(40) *Se=dang kiccil hatthu kaasi-la*
1S= DAT small INDEF give-IMP
'Please give me a small one.' (K081112eli01)

3.3.4 Nominal deixis

Just like the indefiniteness marker, the deictics can also occur with both nominals and Flexible Adjectives, as seen in example (41).

(41) *Se=dang ini kiccil=yang kaasi*
1S=DAT PROX small= ACC give
'Give me the small ones.' (K081112eli01)

3.3.5 Nominal number

The last morpheme we had analyzed as indicative of membership in the noun class was the plural marker *pada*, which can also combine with Flexible Adjectives, as in (42).

(42) *Sedang ini kiccil pada kaasi*
1S= DAT PROX small PL give
'Give me those small ones.' (K081112eli01)

We have seen up to now that Flexible Adjectives can take the morphology we introduced as verbal as well as the morphology introduced as nominal. This entails that we have to reformulate our observations about the cues these features provide: If the morphosyntactic cues exclude a lexeme from the noun class, we are dealing with a verb; if they exclude it from the verb class, we are dealing with a nominal. If no exclusion cues exist, we are dealing with a Flexible Adjective.

[9] An anonymous review asks: 'Does this only take place when there is a lexical gap, or can a converted adjective substitute for an existing verb with comparable meaning? Are there any adjectives at all that do not convert, and if so, is this strictly a function of the semantics of the adjective? Can exceptional adjectives be grouped in well-definable semantic classes? Are borrowed lexical items converted as easily as etymologically Malay items?' In this paper, I cannot survey each and every SLM adjective. As far as I can tell, conversion can take place with any adjective. I have not found any Adjective that could not convert to a verb (see Gil (this volume) for methodological arguments about the burden of proof). As a consequence, classification of the exceptions is not possible: there are no exceptions. Furthermore, there is no indication of conversion being ruled out when a verb with comparable semantics exists. Finally, conversion can also be found with loanwords like 'late'.

3.3.6 Negation pattern

From the discussion so far it appears that there is not a single distributional property that uniquely identifies members of the class of Flexible Adjectives. Flexible Adjectives are characterized by the possibility to behave as nouns or verbs, as the case may be. We have to amend this conclusion slightly, as there is one characteristic that irrefutably indicates membership in the Flexible Adjectives class: the negation pattern. While verbs are normally negated with a prefix, and nouns by *bukang*, most Flexible Adjectives are negated by postposed *thraa* (46) and (47).

(46) *Se thiiňggi thraa*
I tall NEG
'I am/was not tall.' (K081112eli01)

(47) *Samma bannyak responsible thraa*
every much responsible NEG
'They are all not very responsible.' (B060115cvs01)

When *thraa* negates a Flexible Adjective, it carries no indication of tense, as examples (46) and (47) show. *Thraa* can also be used with nouns, but then indicates absence as in example (48). This example could not possibly mean 'It is not a Malay political party'.

(48) *Malay political party atthu thraa*
Malay political party INDEF NEG.EXIST
'There is no Malay political party.' (K051206nar12)

Furthermore, *thraa* can also be used with verbs, but then indicates the perfect tense, as in (49).

(49) *Kithang baaye mlaayu arà-oomong katha incayang biilang thraa*
1PL good Malay NONPAST-speak QUOT NEG 3S say NEG.PERF
'He has not said that we speak good Malay.' (B060115prs15)

The functions of *thraa* are thus different depending on whether it is used in combination with a noun, a verb, or a Flexible Adjective. This can then be used as a shortcut to establish class-membership. Some Flexible Adjectives have a divergent negation pattern where *thàrà-* can be used for all tenses (Slomanson 2006). Because of this temporal vagueness, they can be distinguished from verbs, where *thàrà-* always indicates past tense.

The two different classes of Flexible Adjectives do not show obvious phonological, morphological or semantic characteristics, and different speakers assign different lexemes to them. The different negation patterns thus provide a means to quickly check the category membership of a certain lexeme (Table 1).

	Past	Perfect	Present	Future
Verb	*thàrà*-V	V *thraa*	*thama*-V	
Noun	*bukang*			*bukang* *thama-jaadi*
FlxAdj1	FlxAdj *thraa*			FlxAdj *thraa* *thama*-FlxAdj [10]
FlxAdj2	*Thàrà*-FlxAdj			FlxAdj *thraa* *thama*-FlxAdj

Table 1. Negation as a cue for word class membership.[11]

This particular negation pattern is changed, however, when a Flexible Adjective is converted to a noun or a verb. In that case, the normal verbal or nominal negators are used, as described above. As an illustration, the unconverted Flexible Adjective *kaaya* 'rich' denotes a state and is normally negated by *thraa* (50).

(50) *Se* ***kaaya*** ***thraa***
1S rich NEG
'I am/was/will not (be) rich.' (K081104eli06)

When this lexeme converts to a verb, it denotes a process ('becoming rich'). In this case, it is negated with the verbal negator *thàrà-* for past reference (51) or *thama-* for non-past reference. Note that (51) is not temporally vague but has necessarily reference to the past.

(51) *Se* ***thàrà-kaaya***
1S NEG.PAST[verbal]-rich
'I did not become rich' (K081104eli06)

This conversion can occasionally be found with past and present time references. It is close to obligatory with predications with future time reference. In that case, the conveyed message is not 'will not be PROPERTY' but rather 'will not become PROPERTY' (see example (52)).

(52) *Inni pukuran=yang mà-gijja* ***thamau-gampang***
PROX work= ACC INF-make NEG.NONPAST-easy
'To do that kind of work will never be(come) easy.' (K081106eli01)

If this 'change-of-state'-reading is not intended, the normal negation pattern can be used. This is much less common. Example (53) illustrates this for Type 1 Flexible Adjectives.

(53) *Incayang=pe dudukan hathiyan thaaun=ka=le laile* ***thàrà-baae***
3S =poss behaviour other year= LOC=ADDIT still NEG-good
'His behaviour will still not be good (=remain bad) even next year.' (K081104eli06)

[10] Negated adjectival predicates with future time reference are normally negated as converted verbs with the meaning 'will not become the case' rather than as Flexible Adjectives with the meaning 'will not be the case'.

[11] Negation can also be used to establish subclasses of *Flexible Adjectives* (FlxAdj1 and FlxAdj2), but these have no further properties differentiating them.

Similarly to verbal conversion discussed above in (51), *kaaya* 'rich' can be converted to a noun and then be used referentially. In this case, it is negated with the nominal negator *bukang*.

(54) *Se* ***kaaya*** ***bukang****;* *incayang=jo* *kaaya*
1s rich NEG.NONV 3S=EMPH rich
'I am not the rich person, the rich person is him.' (K081104eli06)

Membership in the class of Flexible Adjectives can thus be established by the negation with postposed *thraa*, but the other negators can also be found following conversion to a verb or a noun, as the case may be.

3.3.7 Negative features

While there is morphology whose presence can serve to exclude a candidate from the class of verbs (e.g. the plural marker) or from the class of nouns (e.g. irrealis prefix), no such features exist for Flexible Adjectives.

3.4 Summary

SLM lexemes can be categorized according to distributional criteria. There is morphology that cannot combine with nouns, and there is other morphology that cannot combine with verbs. The occurrence of one of these morphemes rules out membership in the respective class. Flexible Adjectives can take any morphology. This is summarized in Table 2.

	Noun	Verbs	Flexible Adjectives
Verbal prefixes	-	+	+
Nominalizer	-	+	+
Indefininiteness	+	-	+
Deictic	+	-	+
Plural	+	-	+
Thàrà/thama-	-	+	+
Bukang	+	-	+
Predicative present tense negation with *thraa*	-	-	(+)

Table 2. Overview of defining characteristics of lexical categories (notice that predicate negation with *thraa* is only possible for some Flexible Adjectives)

4 Semantic correlates of the word classes

On a pretheoretical level, adjectives encode properties as their core meaning (Sasse 1993, Sassse 1995, Croft 2001). This core meaning is also found in the following two cases for SLM: modification (55) and property assignment (56).

(55) *Incayang [atthu bìssar orang]*
3S INDEF big man
'He is a big man.' (K081112eli01)

(56) *Aanak thiiňggi*
child tall
'The child is tall.' (K081103eli02)

When Flexible Adjectives are converted to nouns (i.e. they co-occur in a positively nominal context), they encode objects/individuals (cf. Wierzbicka 1986).

(57) *Incayang [kitham=pe bìssar]*
3S.POLITE 1PL=POSS big
'He is our boss.'

When Flexible Adjectives are converted to verbs, (i.e. they co-occur in a positively verbal context), they denote processes.[12]

(58) *Incayang arà-bìssar*
he NONPAST-big
'He is growing up.'

We can summarize this by saying that Flexible Adjectives adapt their semantics to the nominal/object, verbal/process or 'adjectival'/property context that the morphosyntactic frame might require. This means that three cases can be distinguished: [1] in a nominal context, they denote objects, [2] in a verbal context, they denote processes, and [3] elsewhere, they denote properties. This regular shift of semantics makes SLM a member of type (ii) flexibility (Haig 2006).

The distributional behaviour of Flexible Adjectives is then best captured in the structuralist approach by assuming conversion/zero-derivation.

(59) FlxAdj+∅ = N

(60) FlxAdj+∅ = V

Since the semantic shift that Flexible Adjectives undergo under conversion is completely regular, this can be analyzed as word-level-conversion (Don and van Lier, this volume). There are a handful of lexemes that can function both as nouns and verbs in SLM (not Flexible Adjectives), but the semantic relation between them is arbitrary and not regular. Examples for these are *nyaanyi* 'sing/song' and *jaalang* 'walk(V)/street'. These pairs are then instances of root-level-conversion in the model Don and van Lier propose.

[12] Similar semantic changes are attested for Mauritian Creole (Alleyne 2000: 131) and Kharia (Peterson, this volume).

5 Functional approach

The structuralist distributional approach works for individual languages, but is impossible to apply when comparing languages since the criteria to establish the categories are dependent on every individual language (Haspelmath 2007). To overcome the limited usefulness of the morphosyntactic approach, Kees Hengeveld and colleagues have developed a classification of parts-of-speech systems that is based on discourse-pragmatic functions rather than on morphosyntactic criteria (Hengeveld 1992, Hengeveld et al. 2004, Hengeveld and van Lier 2008, Hengeveld and Mackenzie 2008). While distributional criteria cannot be used to identify word class membership across languages, it is assumed that discourse pragmatic criteria can be used for this purpose. Languages do not all employ the same set of formal (morphosyntactic) categories such as 'plural marker' or 'case marker', but the discourse notions PREDICATION, REFERENCE and MODIFICATION are probably relevant for all languages (cf. Croft 2001: 84-5,87). In fact, Hengeveld and colleagues use the category labels PREDICATE, REFERENTIAL PHRASE, HEAD and MODIFIER to arrive at the following four discourse functions (Table 3):

	head	modifier
predicate phrase	head of Predicate Phrase	modifier of Predicate Phrase
referential phrase	head of Referential Phrase	modifier of Referential Phrase

Table 3. Discourse-pragmatic phrasal functions according to Hengeveld

Each lexeme can be tested as to whether it can be used in one of these four discourse functions as it is, or whether additional material like a copula is needed if the lexeme is to occur in this function. Hengeveld (this volume) gives the following illustration:

> The four functions and their lexical expression can be illustrated by means of the English sentence in (61).[13]
>
> (61) *The tall*$_{A}$ *girl*$_{N}$ *sings*$_{V}$ *beautifully*$_{MAdv}$
>
> English can be said to display separate lexeme classes of verbs, nouns, adjectives and (derived) manner adverbs, on the basis of the distribution of these classes across the four functions identified in Figure 1: verbs like sing are used as heads of predicate phrases; nouns like girl as heads of referential phrases; adjectives like 'tall' as modifiers in referential phrases; and manner adverbs like 'beautifully' as modifiers in predicate phrases. Crucially, none of the content lexemes in (1) could be used directly in another function, i.e. without morpho-syntactic adaptation. Thus, in this example there is a one-to-one relation between function and lexeme class. Parts-of-speech systems of this type are called differentiated, and the lexical classes can all be said to be specialized for a certain propositional function.

[13] Notice that we use our numbering here, i.e. not Hengeveld's.

There are other parts-of-speech systems in which there is no one-to-one relation between the four propositional functions identified and the lexeme classes available. These systems are of two types. In the first type, a single class of lexemes is used in more than one propositional function. Such lexeme classes, and the parts-of-speech systems in which they appear, are called ‘flexible’. In the second type, called 'rigid', some propositional functions have no lexemes associated with them.

In order to arrive at word classes, we can now analyze every lexeme in a given language to see which of these functions it can fulfil without further measures (like copula support) being taken. Some lexemes can fulfil only one function, while others can fulfil more than one. Lexemes that can fulfil exactly the same functions are said to be in the same class. Table 4 lists 14 classes and the functions that can be fulfilled by the members.

	HP	HR	MR	MP
Verb	+	-	-	-
Noun	±	+	-	-
Adjective	±	-	+	-
Manner adverb	-	-	-	+
Predicative	+	-	-	+
Nominal	-	+	+	-
Modifier	-	-	+	+
Non-verb	-	+	+	+
Contentive	+	+	+	+

Table 4. Taxonomy of parts of speech as defined by discourse-pragmatic functions (adapted from Smit 2006).[14]

Table 4 contains familiar terms from parts-of-speech research, like noun or verb. A verb is defined in this theory as a lexeme that can be used as head of predicate phrase as it is, but must undergo further measures to be used in other functions. A manner adverb is defined as a lexeme which can only be used as modifier of predicate without further measures being taken, but cannot be used in other functions as is. Nouns are specialized for the function head of referent and adjectives are specialized for the function modifier of referent. Both may or may not be used as head of predicate. Parts of speech with only one function are called rigid.

These are the basic four well-known parts of speech reinterpreted in a discourse framework. Parts of speech which have more than one function are called flexible and are given further down the table, like the nominal, which can be used as head or modifier of referent, but nowhere else; the non-verb, which can be used anywhere but as the head of predicate; and finally the Contentive, which excels in all functions.

In this introduction, I have outlined the functional definition of parts of speech and the resulting taxonomy of parts of speech as well as the parts-of-speech constellations that are said to exist. In the following, I will apply these definitions to the Sri Lanka Malay distributional classes established above and see which parts of speech can be found in this language.

[14] ‘Noun’ and ‘Adjective’ are different from the other classes in that they have ± for the use as head of Predicate Phrase. They may or may not require further measures to be used in this function. Other combinations of discourse-pragmatic functions (e.g. + - + -) are logically possible, but have not been proposed up to now.

5.1 The verb

Members of the SLM verb class established above can be used as the head of predicate without further measures being taken. They cannot be used for any other function as they are (cf. Table 5). This means that this class is indeed also a verb in the Hengeveldian sense. We will illustrate this in the following sections.

	Head	Modifier
Predicate phrase	no further measure *(62) Kithan=le anthi-banthu* 1PL ASSOC= IRR-help 'We will help, too.' (B060115nar02)	further measure: reduplication . *(63) Kancil* ***lompath˜lompath*** *arà-laari* rabbit jump˜jump NONPAST-run 'The rabbit runs away jumping.' (K081112eli01)
Referential phrase	further measure: infinitive formation (64) ***Mà****-banthu baae* INF -help good 'To help is good.' (K081112eli01)	further measure: relative clause formation *(65)* [*Arà-maakang*] *oorang* NONPAST-eat man 'The eating man.' (K081112eli01)

Table 5. The SLM verb in different discourse functions.

Example (62) shows that lexemes of the class of verbs can be used as Head of Predicate Phrase without further measures being taken. For the other functions, further measures are necessary. In order to be used as modifiers in the predicate phrase, verbs have to be reduplicated as shown in (63). In order for verbs to head a referential phrase, the infinitive prefix *mà-* has to be used (64).[15] For modifier function in the referential phrase, a relative clause construction must be employed (65).

5.2 The noun

Members of the SLM noun class can be used as head or modifier of the referential phrase. They cannot be used as head or modifier of the predicate phrase (cf. Table 6).

Example (68) shows the use of the lexemes *baapa* 'father' and *ummas* 'gold' as heads of the referential phrase. No further measures are required for these lexemes to be used in this position. Lexemes of the SLM noun class can be used to modify a head of referential phrase, like *komplok* 'bush', which modifies *pohong* 'tree' in (69), and is further modified by *rooja kumbong* 'rose flower', itself consisting of the head *kumbang* 'flower' and the modifier *rooja* 'rose'. Lexemes of the noun class in SLM can also be used as head of predicate phrase without having to resort to further measures such as the use of a copula, as shown in (66). In order to be used as a modifier of predicate, SLM lexemes of the noun class have to take the dative marker *=nang*. Without it, the sentence in (67) would be ungrammatical; hence it is a further measure.

[15] An alternative is the nominalizer *–an.*

	Head	Modifier
Predicate phrase	no further measure *(66) Sindbad the Sailor ∅ hatthu Muslim, mlaayu bukang* Sindbad the Sailor INDEF muslim, Malay NEG.NONV 'Sindbad the Sailor was a Moor, he was not a Malay.' (K060103nar01)	further measure: adverbialization *(67) Incayang* ***swaara=nang*** *arà-oomong* he sound= DAT NONPAST-speak 'He speaks with a loud voice.' (eli14122005) (K081112eli01)
Referential phrase	no further measure *(68) Se=ppe baapa incayang=nang ummas su-kaasi* 1S=POSS father NONPAST 3S= DAT gold PAST -give 'My father gave him the gold.' (K070000wrt04)	no further measure (69) *Panthas rooja kumbang pohong komplok duuwa asà-jaadi su-aada* beautiful rose flower tree bush two CP-grow PAST-exist 'Two beautiful rose bushes had grown.' (K070000wrt04)

Table 6. The SLM noun in different discourse functions.

5.3 The Flexible Adjective

We have established that the SLM verb is a rigid class, and that the SLM noun is moderately flexible. We now turn to SLM Flexible Adjectives, which can be used for any discourse function. Table 7 shows that Flexible Adjectives can be used in all four discourse functions without further measures being taken. This means that SLM Flexible Adjectives are 'Contentives' in the Hengeveldian terminology.

	Head	Modifier
Predicate phrase	no further measure *(70) Samma oorang baae* all man good 'All men are good.' (B060115cvs13)	no further measure *(71) Incayang pullang arà-oomong* he slow NONPAST -speak 'He speaks softly.' (eli14122005) (K081112eli01)
Referential phrase	no further measure *(72) Incayang hatthu iitham* 3S INDEF black 'He is a dark person.' (K081103eli02)	no further measure *(73) Se=ppe bìssar lae ruuma aada* 1S=POSS big other house EXIST 'There is another big house of mine.' (B060115cvs09)

Table 7. The SLM Flexible Adjective in different discourse functions.

Members of the SLM Flexible Adjective class can be used as head of term, as (72) shows. *Iitham* 'black' is a Contentive, which is used as the head of the referential phrase. In example (73), *bìssar* 'big' modifies the head *ruuma* 'house', without further measures being taken. Predicate phrases can have a Flexible Adjective in modifier position without further measures being taken: example (71) shows the use of the Flexible Adjective *pullam* 'slow' to modify the head verb *oomong* 'speak'. Note that this contrasts with the obligatory use of *=nang* in (67) for nouns in this position.

A difference between Flexible Adjectives and Verbs is that the former can be used in different predication types. The Property Assignment Predication is exemplified by (70) and

(74). No additional material is needed for a Flexible Adjective to occur in this predication type.

(74) *Aanak thiiňggi*
child tall
'The child is tall.' (K081103eli02)

Another predication type where Flexible Adjectives can occur in is the Dynamic Predication, signaled by TAM-morphology on the predicate. Members of the SLM Flexible Adjective class can occur in this position without further measures being taken, as (75) shows.

(75) *Itthu=nam blaakang=jo, kitham pada* ***anà-bìssar***
DIST after= EMPH 1PL PL PAST-big
'After that, we grew up.' (K060108nar02)

Note that aktionsart is static in (70) and (74), typical for the Property Predication in SLM, while (75) is an accomplishment, indicated by the use of the Dynamic Predication.
The following example shows one and the same lexeme *dhiinging* 'cold' used in both predication types:

(76) a *Thee dhiinging*
tea cold
N ADJ
'The tea is cold.' (K081112eli01)

b *Thee arà-dhiinging*
tea NONPAST-cold
N V
'The tea is getting cold.' (K081112eli01)

This is irrelevant for the Hengeveldian approach, which does not distinguish between different types of predications; nevertheless a more fine-grained approach could be warranted (Haig 2008). The third predication type a Contentive can occur in is the Class Membership Predication, which uses the indefiniteness marker *atthu*. This is not very frequent. An example is given in (78).

(77) *Incayang hatthu iitham*
3s INDEF black
'He is a dark person.' (K081103eli02)

5.4 Summary

We have seen examples of the use of members of different distributional classes for different discourse functions. Table 8 sums up the findings of the discourse functions the SLM parts of speech can fulfill and gives the names in Hengeveldian taxonomy: The SLM Verb can only be used as head of predicate, and is therefore a verb (Table 4). The SLM Noun can be used as a head or modifier of the referential phrase and is therefore a nominal (Table 4). The SLM Flexible Adjective can be used for any function and is therefore a Contentive (Table 4).

Head of predicate phrase	Head of referential phrase	Modifier of referential phrase	Modifier of predicate phrase
SLM Verb (verb)	SLM Noun (nominal)		
SLM Flexible Adjective (contentive)			

Table 8. Schematic representation of the SLM parts of speech and propositional functions

The functional approach to parts of speech started out with the premise of cross-linguistic comparability and chose to use discourse functions instead of distribution classes to establish parts of speech. This procedure could be successfully applied for the redefinition of SLM parts of speech.

6 Diachrony of PoS systems

6.1 Specialization and generalization

Parts-of-speech systems are very often seen as static. This makes it easy to deal with them, but on the other hand we know that languages are not static in that there is always some part of the grammar that is changing. We know that languages change their word order, their morphological type, their alignment type, their stress system etc. Parts-of-speech systems can also change over time.[16][17]

An adequate theory of parts of speech must then be able to capture the transitions from one type to another. Basically, we see can two historical developments: loss of discourse functions and gain of discourse functions (Table 9).[18]

	FA		FB
Lexeme at t_0	+		+
	↓	↙	
Lexeme at t_1	+		-

Loss of a function, specialization. The lexeme has both discourse functions FA and FB at t_0, but has lost FB at t_1.

	FA		FB
Lexeme at t_0	+		-
	↓	↘	
Lexeme at t_1	+		+

Gain of a function, generalization. The lexeme has only function FA at t_0, but has gained FB at t_1.

Table 9. Direction of change in parts-of-speech systems.

Linguistic change is gradual. This means that not all of the lexemes will be able to be used in a new function at once. Rather, there will be pioneers and latecomers. Pioneers are normally derived lexemes, while latecomers are lexemes in closed classes, a kind of residue. Derived classes are then the first sign of incipient specialization, while closed classes indicate the last stages of generalization. We will take a look at specialization and leave the processes leading to generalization as a future research project.

[16] To be in constant change is actually seen as the normal case for parts-of-speech systems in Hengeveld (1992: 69).
[17] Cf. the flexibility of Late Archaic Chinese, which gave way to the more rigid modern system (Bisang 2008b).
[18] Vogel (2000) discusses the emergence of lexical categories, which she calls grammaticalization and the disappearance thereof, which she terms degrammaticalization. The former is similar to specialization, and the latter to generalization as used in this paper. Vogel's use of the terms applies to the emergence/disappearance of differentiated parts-of-speech, not to the gain/loss of discourse functions, although the two are obviously related.

6.2 Diachrony of the SLM system

Sri Lanka Malay is the descendant of varieties of Trade Malay spoken in Jakarta.[19] While we have no description of the parts-of-speech system of this historic variety, two factors make it seem very likely that it was a language with an extremely flexible part-of-speech system: genetic affiliation to Malayic and Austronesian, and its origin as a trade language.
Malay languages (and Austronesian languages in general) are characterized by the fact that there are no clear boundaries between the traditional lexical word classes as recognized in the better-known European languages. Himmelmann (2005: 128f.) writes:

> [On the lexeme level], it is frequently noted in descriptions of western Austronesian languages that lexical bases (roots) are underdetermined in allowing both nominal and verbal derivations or uses. Alternatively, a basic distinction, between nouns and verbs (and possibly adjectives) is made for lexical bases but then it is stated elsewhere in the grammar that nominal bases can be used as (morphosyntactic) verbs essentially in the same way as verbal bases. [...]
> Multifunctional lexical bases, which occur without further affixation in a variety of syntactic functions occur [...] with some frequency in [...] many Malayic varieties.

Himmelmann's description can be reworded as 'lexemes can be used in a variety of functions without further measures being taken', matching precisely the Hengeveldian definition of flexible parts of speech. An example of this flexibility is given by Ewing (2005: 230f.) for Colloquial Indonesian as spoken in Java.

(78) *Kalau saya cerita begini* *COLLOQUIAL INDONESIAN*
if 1S story like.this
'If I tell a story like this.'

(79) *Makan juga nggak boléh di-taroh di situ* *COLLOQUIAL INDONESIAN*
eat also NEG allow UV-put LOC there
'Food also isn't allowed to be put here.'

In (78) the word *cerita* 'story' is used without a verbalizer when it serves as the main predicate (head of predicate phrase). In (79) the word *makan* 'eat' is used without a nominalizer when it serves as the head of a referential phrase.[20] Lexical categories in another variety of Malay, Riau Indonesian, have been investigated inseveral papers by Gil (for example, Gil 1994), who claims that Riau Indonesian lacks distinctive word classes.[21] The

---

[19] Not to be confused with Betawi Malay, spoken as a native language by Jakarta Malays at the same period, and also distinct from Jakartan Indonesian as spoken today. Jakarta Malay as described by Grijns (1991) is also an offshoot of these varieties. See also Paauw (2004) for a sociohistorical and linguistic analysis of the origin of the SLM immigrants and Adelaar (1991) and Adelaar and Prentice (1996) for a different view. Both are discussed in Nordhoff (2009).
[20] Similar ideas are expressed by Prentice (1990: 920), who observes that colloquial spoken Indonesian is characterized by extensive elision of affixes, resulting in a greater degree of overlap between word classes.
[21] An anonymous reviewer remarks that claims about the radically monocategorial status of Malay varieties are controversial and that 'there are informal claims that [the existence of semantic constraints] makes the radical claim untenable.' In this paper, I stick to published work and do not take into account informal claims, but I acknowledge their existence. As for methodological issues in Malay linguistics with special consideration of parts-of-speech, see Gil (this volume).

trade language spoken in Jakarta before the mercenaries left for Sri Lanka evolved thus in an ecology characterized by languages with weak or absent distinctions between word classes. If we add to this that the Malay variety the mercenaries communicated in was not their mother tongue, the lack of morphologically distinct word classes becomes even more likely given that trade languages (pidgins, lingue franche) tend to be morphologically more reduced than the related language spoken by natives. It is extremely likely that the cocktail of colloquial Malay varieties without specialized word classes distilled into a morphologically reduced trade language would result in a language without distinctive word classes. The same analysis has been proposed in Paauw (2004):

> In [Trade Malay], a word can function variously as a noun, verb, adjective or function word without any change in form, depending entirely upon the context and word order for its function.

Having established that the ancestor of SLM was most likely a monocategorial language, we must ask ourselves the question why SLM is no longer monocategorial today. The equivalents of examples (78) and (79) above in SLM are given below. We see that *criitha* 'story' cannot be used as a transitive predicate in SLM in (80). The use of the transitive verb *biilang* 'say' is obligatory. Example (81) illustrates that the use of the nominalizer *-an* is obligatory in SLM. Leaving it out would result in ungrammaticality.

(80) *See giini hatthu criitha=ke kala-**biilang***
1S like.this NDEF story= SIMIL if-say
'If I tell a story like this.' (K081112eli01)

(81) *Makan-**an**=le siithu mà-thaaro thàràboole*
eat-NMLZR=ASSOC there INF-put cannot
'Food also isn't allowed to be put here.' (K081112eli01)

The reasons for the creation of distributional restriction on the use of nominals and verbs in certain syntactic slots can be sought on two different grounds: language contact on the one hand, and cognitive factors relating to time-stability on the other hand. These will be explored in the following two sections.

6.3 Language contact

SLM has been in intimate contact with Sri Lanka's majority languages Sinhala and Tamil for at least 300 years, and its structure has been heavily influenced by the adstrates (Adelaar 1991, Smith et al. 2004, Slomanson 2006, Ansaldo 2008, Nordhoff 2009). Different authors have argued for Tamil (Smith 2003, Smith and Paauw 2006, Smith et al. 2004) or Sinhala (Ansaldo 2008) being the more important contact language.

This is difficult to evaluate, since Sinhala and Tamil share a very similar typological markup (Smith 2003). Smith (2003) argued that SLM follows Tamil where Sinhala and Tamil diverge. These claims were critically reassessed in Nordhoff (2009), who shows that the data do not warrant this conclusion. Smith's data could all be shown to not point toward clear Tamil influence. They mostly point towards joint influence, with some phenomena only being accountable in terms of Sinhala influence. In the present paper, I do not take a stance as to whether Sinhala or Tamil is the more important contact language. Following Nordhoff (2009), I take this to be an empirical question which is not answered yet.

As it happens, Sinhala and Tamil are both rigid languages, i.e. languages which make a distinction between nouns, verbs and adjectives.[22] Sinhala has a large open class of adjectives (Gair 1967: 21), while Tamil also has adjectives, but the class is closed and hosts only a handful of underived lexemes (Schiffman 1999: 125). In both languages, these classes are usually designated as 'adjectives' and mainly host lexemes denoting property concepts. The term 'adjective' is used e.g. in Gair (1967) or Lehmann (1989), which corresponds in this case to Hengeveldian terminology, i.e. they can only be used as modifier of a Referential Phrase without further measures being taken. Sinhala and Tamil both have a class of verbs, which can only be used in predicative function. In order to use verbs in referential function, nominalizations have to be used in Sinhala (Garusinghe 1962: 62ff) and Tamil (Lehmann 1989: 300ff). Nouns can be used both referentially and predicatively in both languages.[23]

An anonymous reviewer remarks that it is common for Sinhala nouns to appear in 'verbal' (= predicative) position without any additional morphosyntactic marking. This is true (for Tamil as well, actually), but does not present a problem in the Hengeveldian approach, since the function HEAD OF PREDICATE PHRASE has a ± for nouns (see Table 4). Nouns may or may not take additional morphology in order to appear in predicative position. What distinguishes nouns from verbs is that verbs must not be able to appear in referential position without further measures being taken. This is the case in Sinhala, where the nominalizers *-ma* or *eka* must be used (Garusinghe 1962: 62ff), and in Tamil, where *-al*, *-ttal*, *-gai* or *-adu* are used (Lehmann 1989: 300ff).

The grammar of SLM has been influenced by its adstrates in all the main areas: phonology (retroflex consonants; Bichsel-Stettler 1989, Tapovanaye 1995), morphology (cases, finiteness distinctions; Smith et al. 2004, Ansaldo 2008, Slomanson 2006), syntax (SOV word order; Adelaar 1991) and semantics (e.g. the modal systems; Slomanson 2006).

Apparently, language contact has also affected the SLM parts-of-speech system. SLM has converged to match the overall structure of the island's languages and become part of the Sri Lankan sprachbund (Bakker 2006, Ansaldo 2008). In linguistic areas characterized by an absence of rigid parts of speech like large parts of Indonesia (Gil 2001), other Malay varieties could keep their lexical flexibility, but this was not possible in Sri Lanka. While these external conditions trigger the language change in the mind of the individual speaker, there must be some cognitive disposition in the minds of the speakers, which allows them to assign lexemes to lexical categories. In the following section I will argue that [±TIME-STABILITY] is the determining factor for the SLM development.

6.4 The cognitive factor: time stability

The specialization of formerly underdetermined lexemes in the history of SLM happened along the semantic fault lines of [±TIME-STABILITY] (Givón 1984: 51): words typically used for objects (time-stable entities) lost the ability to be used as (dynamic) predicates and were

---

[22] At least on the word level. On the phrase level, class membership is less clear cut, at least in Sinhala cf. Gair (2003: 796).

[23] Besides the classical use for nominal predicates (class-inclusion), there are some additional event predicates which are coded by nouns in Sinhala (action nominals in Gair and Paolillo 1989). An anonymous reviewer points out that this Sinhala structure is not found in SLM. This is true. The reviewer continues that Sinhala influence is therefore unlikely. This objection does not hold. It is possible that Sinhala had an influence on the general configuration of SLM parts-of-speech without that influence extending into very particular predicate types. A wholesale copy of the Sinhala system is no requirement for the postulation of partial influence. The SLM system has moved towards Sinhala and Tamil, but very particular structures of Sinhala (action nominals) or Tamil (closed class of adjectives) are not found in SLM. In spite of this, the SLM system is still closer to Sinhala and/or Tamil than to the systems of other Malays varieties.

grouped in the class of nominals, while words normally used in connection with events (entities that are not time-stable) lost their ability to be used for things and were grouped in the class of verbs.

Aayang 'chicken' is an example of a word that usually denotes a concrete object (a time-stable entity) that can only be used referentially in SLM without further measures being taken, while *maakang* 'eat' is an example of a word that most frequently denotes an activity and cannot be used referentially any more in SLM (cf. (81)). In distinction to, say, Riau Indonesian, morphological operations like infinitive prefixes, nominalizers, reduplication etc. are required in SLM in order to use *maakang* in any other function than head of predicate (cf. Table 5 above).

It is well known that semantic features of lexemes and potential pragmatic functions correlate (Croft 1991, 2001; Sasse 1993). Objects ([+time-stable]) correlate with reference, actions ([-time-stable]) correlate with predication and properties (intermediate) correlate with modification. In the case of the diachronic development of SLM parts of speech, words primarily used for [+time-stable] entities (objects) were restricted to their prototypical pragmatic function of reference and barred from other discourse functions. Words primarily used for [-time-stable] entities (events) in turn were restricted to their prototypical pragmatic function of predication and barred from other discourse functions without further measures being taken. The interesting cases are concepts which are intermediate on the scale of time-stability, i.e. properties. For these, there was no clear semantic cue on the time-stability scale as to their prototypically associated discourse function, and hence specialization did not take place: these lexemes can still be used for every discourse function and underlie no restrictions. Being on the middle ground of the time-stability scale helped them retain the lexical flexibility typical of their ancestor variety.

At this point of the analysis, we recapitulate that a certain semantic feature of a certain lexeme evokes a certain semantic class which in turn attracts a prototypical discourse function. Over time, the lexeme becomes intimately connected with the discourse function, and use in non-prototypical functions has to be signaled (relative clause formation, infinitive prefixes and the like).

Put like this, it follows that specialization of word-classes is the morphosyntactic reflex of cognitive association of lexemes with prototypical discourse functions. The emergence of new word classes therefore has its origins not in morphosyntax nor in discourse proper, but in semantics and its association to discourse pragmatics.

This association between semantics and discourse function is where language contact kicks in: the differentiated associations between words exclusively used for time-stable entities found in the adstrates (and other words for [-time-stable] entities) had to be mastered by the arriving Malays, who learned the adstrates and applied the newly acquired association to their own language as well. They started with the most obvious cases, those which were clearly [+time-stable] or clearly [-time-stable].

The road for this specialization might actually have been paved beforehand by the existence of optional derivational morphology.[24] The nominalizer *-an* for instance is an affix which survived the state of trade language and is used productively today. There is no evidence whether or not it was obligatory in the variety of the mercenaries. Its use might have been optional, as is the case today in Colloquial Spoken Indonesian (cf. (79)). Speakers would then have the possibility to mark restriction to certain discourse functions if they deemed

[24] See Smit (2006) for similar ideas in a slightly different approach. Smit assumes that derivation is the source of the systems called intermediate in Hengeveld (1992). None of the intermediate systems Hengeveld (1992) proposes fits the SLM data, and the idea of intermediate systems has been abandoned by him since then, so that Smit's approach cannot be applied to the SLM case.

necessary.[25] The existence of a derivational target function indicates that the speakers could indicate this function if so desired, or leave the identification of the discourse function to the hearer if they did not judge the disambiguation necessary. As stated above, the lexemes derived with one of the optional affixes would then be the pioneers in the new lexical categories, while other lexemes slowly follow suit as their association with the particular discourse function grows tighter. Due to the lack of historical records, the precise development is difficult to reconstruct, but we can assume that the ancestor variety was monocategorial, and we know that the variety of today is not. What is missing is the 'pioneer' stage of derived lexemes claiming the new discourse function. This stage is not attested in the development of SLM, but we can find comparable transitory systems in other languages, which I will briefly illustrate in the next section.

6.5 Comparison with other flexible languages

In this section, I will take a look at languages with a class of Contentives and at least one other lexeme class. In Hengeveld and van Lier (2008), we find Kambera (based on Klamer 1998), Santali (based on Neukom 2001) and Samoan (based on Mosel and Hovdhaugen 1992), which all fulfill this criterion. The parts-of-speech constellation of these languages is given in Table 10.

Samoan			
Head of predicate phrase	Head of referential phrase	Modifier of referential phrase	Modifier of predicate phrase
Contentive$_O$			
Verb$_D$			

Kambera			
Head of predicate phrase	Head of referential phrase	Modifier of referential phrase	Modifier of predicate phrase
Contentive$_O$			
Verb$_O$			Manner Adverb$_C$

Santali			
Head of predicate phrase	Head of referential phrase	Modifier of referential phrase	Modifier of predicate phrase
Contentive$_O$			
Verb$_O$	Noun$_D$		

Sri Lanka Malay			
Head of predicate phrase	Head of referential phrase	Modifier of referential phrase	Modifier of predicate phrase
Contentive$_O$			
Verb$_O$	Nominal$_O$		

Table 10. Parts-of-speech systems in various languages with a class of Contentives.[26]

[25] Cf. Marchand (1969) on the 'categorizing' function of derivation. I would like to thank Geoffrey Haig for pointing out this reference to me.

[26] Subscript D signals that a class only contains derived lexemes; subscript C indicates a closed class and subscript O an open class.

We observe that in Samoan, it is possible to derive verbs, but that these derived lexemes are the only ones restricted in their discourse function to head of predication. This is different from Kambera, where underived lexemes can also be found in the class of verbs.[27] In Santali finally, the class of verbs is well established, and yet other derived lexemes specialize for head of Referential Phrase, i.e. they are members of the noun class. The SLM system is added in a fourth table, as a system where both classes of verbs and nominals have progressed beyond hosting only derived lexemes.

Samoan, Kambera, Santali and SLM then illustrate different stages in the specialization of parts-of-speech systems (cf. Table 9). New discourse functions are explored by derived lexemes (verbs in Samoan, nouns in Santali), and later populated by underived lexemes (verbs in Kambera, nominals in SLM). From these four examples, it appears that the development of new lexical categories in this manner proceeds from left to right (or from verbs to nouns), but investigation of more languages, including the diachrony of their parts-of-speech system will be needed to test this hypothesis (cf. Hengeveld and van Lier (2008), who claim that all constraints on PoS-classification converge towards the specialization of verbs, before any other word class.).

7 Conclusion

This paper has shown that Sri Lanka Malay has a maximally flexible class of adjectives, which would be treated as Contentives in Hengeveld's terminology. This maximally flexible class co-exists with the more rigid classes of nominals and verbs, showing that lexeme flexibility and monocategoriality do not necessarily accompany each other, but are orthogonal dimensions.

In diachronic perspective, nominals and verbs evolved from Contentives having lost some discourse functions and thereby their flexibility. This change towards a more rigid system was probably triggered by the adstrates Sinhala and Tamil, in which the discourse function Predication is strongly associated with events and the function Reference, with objects. The fact that SLM already had the possibility to derive forms for more specific discourse functions probably accelerated the contact-induced changes in the SLM parts-of-speech system.

[27] We do not have anything to say about manner adverbs for now and disregard this class.

Abbreviations

ACC accusative; ASSOC associative; COMPL completive; CP conjunctive participle; DAT dative; DEM demonstrative; DIST distal; EMPH emphatic; INDEF indefinite article; INF infinitive; INTERR interrogative; IRR irrealis; LOC locative; NEG negation; NMLZR nominalizer; NONV nonverbal; PL plural; POSS possessive; PROX proximal; QUOT quotative; UV undergoer voice.

References

Adelaar, K. A. (1991). 'Some notes on the origin of Sri Lankan Malay', in H. Steinhauser (ed.), *Papers in Austronesian linguistics, Pacific Linguistics, vol. 1.* Canberra: The Australian National University, 23-37.

Adelaar, K. A. and Prentice, D. J. (1996). 'Malay: its history, role and spread', in S. A. Wurm, P. Mühlhäusler and D. T. Tryon (eds.), *Atlas of Languages of Intercultural Communication* (Volume II.1). Berlin: Walter de Gruyter, 673-693.

Alleyne, M. C. (2000). 'Opposite processes in creolization', in I. Neumann-Holzshuh and E. W. Schneider (eds.), *Degrees of Restructuring in Creole Languages*. Amsterdam: Benjamins, 125-134.

Ansaldo, U. (2005). 'Typological admixture in Sri Lanka Malay. The case of Kirinda Java'. Unpublished manuscript.

--- (2008). 'Sri Lanka Malay revisited: Genesis and classification', in A. Dwyer, D. Harrison and D. Rood (eds.), '*A world of many voices: Lessons from documented endangered languages*. Amsterdam and Philadelphia: Benjamins, 13-42.

Bakker, P. (2006). 'The Sri Lanka Sprachbund: The newcomers Portuguese and Malay', in Y. Matras, A. McMahon and N. Vincent (eds.), *Linguistic Areas - Convergence in Historical and Typological Perspective*. Houndmills NY: Palgrave Macmillan, 135-159.

Bhat, D. (1994). *The Adjectival Category: Criteria for Differentiation and Identification*. Amsterdam: Benjamins.

Bichsel-Stettler, A. (1989). Aspects of the Sri Lanka Malay community and its language. MA thesis, Universität Bern.

Bisang, W. (2008a). 'Precategoriality and argument structure in Late Archaic Chinese', in J. Leino (ed.), *Constructional Reorganization* (Constructional Approaches to Language 5). Amsterdam and Philadelphia: Benjamins, 55-88.

--- (2008b). 'Precategoriality and syntax-based parts of speech', *Studies in Language* 32-3: 568-589.

Bloomfield, L. (1933). *Language*. London: Allen and Unwin.

Broschart, J. (1997). 'Why Tongan does it diffferently', *Linguistic Typology* 1-2: 123-166.

Croft, W. (1991). *Syntactic Categories and Grammatical Relations: The Cognitive Organization of Information*. Chicago: University of Chicago Press.

--- (2001). *Radical Construction Grammar*. Oxford: Oxford University Press.

--- (2005). 'Word classes, parts of speech, and syntactic argumentation', *Linguistic Typology* 9-3: 431-441.

Ewing, M. C. (2005). 'Colloquial Indonesian', in K. A. Adelaar and N. P. Himmelmann (eds.), *The Austronesian Languages of Asia and Madagascar*. London and New York: Routledge, 227-258.

Gair, J. W. (1967). 'Colloquial Sinhala Inflectional Categories and Parts of Speech', *Indian Linguistics* 27: 32-45. Reprinted in J. W. Gair (1998), *Sinhala and Other South East Asian Languages*. New York and Oxford: Oxford University Press, 13-25.

--- (2003). 'Sinhala', in G. Cardona and D. Jain (eds.), *The Indo-Aryan Languages*, London: Routledge, 766-818.

Gair, J. W. and Paolillo J. C. (1989). 'Sinhala non-verbal sentences and argument structure', *Cornell Working Papers in Linguistics* 8: 39-78. Reprinted in J. W. Gair (1998), *Sinhala and Other South East Asian Languages*. New York and Oxford: Oxford University Press, 87-110.
Garusinghe, D. (1962). *Sinhala, the Spoken Idiom*. München: Hueber.
Gil, D. (1994). 'The Structure of Riau Indonesian', *Nordic Journal of Linguistics* 17: 179-200.
Gil, D. (2001). 'South Asia as a Linguistic Area: A Reassessment'. Paper presented at the *Twenty First South Asian Languages Analysis Roundtable*, Konstanz (Germany), 8 October 2001.
Givón, T. (1984). *Syntax: A Functional-Typological Introduction*. Amsterdam: Benjamins.
Grijns, C. D. (1991). *Jakarta Malay*. PhD dissertation, Universiteit Leiden.
Haspelmath, M. 2007. 'Pre-established categories don't exist: consequences for language description and typology', *Linguistic Typology* 11-1: 119-132.
Hengeveld, K. (1992). *Non-verbal Predication: Theory, Typology, Diachrony* (Functional Grammar Series 15). Berlin - New York: Mouton de Gruyter.
Hengeveld, K. and Mackenzie, L. (2008). *Functional Discourse Grammar: A Typologically-based Theory of Language Structure*. Oxford: Oxford University Press.
Hengeveld, K. and Rijkhoff, J. (2005). 'Mundari as a flexible language', *Linguistic Typology* 9-3: 406-431.
Hengeveld, K. and Lier, E. H. van (2008). 'Parts of speech and dependent clauses in Functional Discourse Grammar', *Studies in Language* 32-3, 753-785.
Hengeveld, K., Rijkhoff, J. and Siewierska, A. (2004). 'Part-of-speech systems and word order', *Journal of Linguistics* 40-3: 527-570.
Himmelmann, N. P. (2005). 'The Austronesian languages of Asia and Madagascar: typological characteristics', in K. A. Adelaar and N. P. Himmelmann (eds.), *The Austronesian Languages of Asia and Madagascar*. London and New York: Routledge, 110-181.
Hussainmiya, B. A. (1990). *Orang Rejimen: The Malays of the Ceylon Rifle Regiment*. Bangi: Penerbit Universiti Kebangsaan Malaysia.
Klamer, M. (1998). *A grammar of Kambera*. Berlin and New York: Mouton de Gruyter.
Lehmann, C. (1989). Language description and general comparative grammar', in G. Graustein and G. Leitner (eds.), *Reference Grammars and Modern Linguistic Theory*. Tübingen: Niemeyer, 133-162.
Marchand, H. (1969). *The Categories and Types of Present-day English Word-formation: a Synchronic-diachronic Approach*. Munich: Beck.
Mosel, U. and Hovdhaugen, E. (1992). *Samoan Reference Grammar* (Instituttet for sammenlignende kulturforskning, Series B: Skrifter, LXXXV). Oslo: Scandinavian University Press.
Neukom, L. (2001). *Santali*. Munich: Lincom.
Nordhoff, S. (2009). *A Grammar of Upcountry Sri Lanka Malay*. PhD dissertation, Universiteit van Amsterdam.
Paauw, S. H. (2004). *A historical analysis of the lexical sources of Sri Lanka Malay*. MA thesis, York University.
Peterson, J. (2005). 'There's a grain of truth in every "myth", or, Why the discussion of lexical classes in Mundari isn't quite over yet', *Linguistic Typology* 9-3: 391-441.
Prentice, D. J. (1990). 'Malay (Indonesian and Malaysian)', in B. Comrie (ed.), *The World's Major Languages*. London: Routledge, 913-935.
Sasse, H.-J. (1993). 'Das Nomen - eine universale Kategorie?' *Sprachtypologie und Universalienforschung* 46: 187-221.

--- (1995). 'Syntactic categories and subcategories', in J. Jacobs et al. (eds.), *Syntax* (Handbücher zur Sprach- und Kommunikationswissenschaft 9.1). Berlin and New York: Mouton de Gruyter, 646-686.
Schiffman, H. (1999). *A Reference Grammar of Spoken Tamil*. Cambridge: Cambridge University Press.
Slomanson, P. (2006). 'Sri Lankan Malay morphosyntax: Lankan or Malay?', in A. Deumert and S. Durrleman (eds.), *Structure and Variation in Language Contact*. Amsterdam: Benjamins, 135-158.
Smit, N. (2006). 'Intermediate PoS-systems - Clues from derivational morphology'. Paper presented at the conference *Universality and Particularity of Parts-of-Speech systems*, Amsterdam, June 2006.
Smith, I. R. (2003). 'The provenance and timing of substrate influence in Sri Lankan Malay; Definiteness, animacy and accusative case marking.' Paper presented at the *South Asian Language Analysis Roundtable XXIII*, Austin, Texas.
Smith, I. R. and Paauw, S. (2006). 'Sri Lanka Malay: Creole or convert', in A. Deumert and S. Durrleman (eds.), *Structure and Variation in Language Contact*. Amsterdam: Benjamins, 159-182.
Smith, I. R., Paauw, S. and Hussainmiya, B. A. (2004). 'Sri Lanka Malay: the state of the art.' *Yearbook of South Asian Languages* 2004: 197-215.
Tapovanaye, S. (1995). *Vowel Length in Sri Lankan Malay*. MA thesis, University of Hawaii.
Vlekke, B. H. M. (1943). *Nusantara: a history of the East Indian Archipelago*. Cambridge: Harvard University Press.
Vogel, P. M. (2000). 'Grammaticalization and part-of-speech systems', in P. M. Vogel and B. Comrie (eds.), *Approaches to the Typology of Word Classes*, (Empirical Approaches to Language Typology 23). Berlin and New York: Mouton de Gruyter, 259-284.
Wierzbicka, A. (1986). 'What's in a noun (or: How do nouns differ in meaning from adjectives?)', *Studies in Language* 10: 353-389.

Word-class systems between flexibility and rigidity: an integrative approach

Walter Bisang

1 Introduction

Definitions of parts of speech are based on the following four prerequisites: (i) semantic criteria, (ii) pragmatic criteria, (iii) formal criteria and (iv) lexicon-syntax mapping (Sasse 1993a, b, Bisang 2008a, 2011). The first two prerequisites are characterized by semantic concepts such as action, object and property and pragmatic functions such as predication, reference and modification (cf. Croft's 1991, 2000, 2001 conceptual space for parts of speech). The formal criteria are provided by the morphosyntactic formats that are used for the expression of the above semantic concepts and pragmatic functions. Finally, lexicon-syntax mapping is relevant for parts of speech because individual lexical items are associated with syntactic slots that correspond to categories such as verb, noun and adjective.

Hengeveld's (1992) classification of parts of speech combines the first two criteria (semantic and pragmatic) and then looks at two parameters that are defined by how lexemes are mapped onto syntax and by the grammatical markers that are involved (also cf. Hengeveld et al. 2004, Hengeveld and Rijkhoff 2005, Rijkhoff 2008). Both parameters are characterized by different degrees of flexibility and can be integrated into a Typology of Flexibility (Van Lier and Rijkhoff, *this volume*). The parameter of lexicon/syntax mapping, i.e. the lexical parameter or L parameter, is based on four different functions that are associated with syntactic slots, i.e., heads of predicate phrases with verbal slots, heads of referential phrases with nominal slots, modifiers in referential phrases with adjectival slots and modifiers in predicate phrases with adverbial slots. In a parts-of-speech system whose lexemes L are flexible, one and the same lexeme can take different slots. Extreme examples of such languages are Tongan (Broschart 1997), Samoan (Mosel and Hovdhaugen 1992) and Late Archaic Chinese (Bisang 2008a, b and section 2). In these languages, individual lexemes are not preclassified in the lexicon for the syntactic slots of verb, noun and adjective, i.e., a lexical item can occur in each of these slots without changing its form (cf. Bisang 2008a on precategoriality). In languages with a more rigid lexical parameter (L parameter), the assignment of individual lexemes to the four syntactic slots is more restricted. The parameter of grammatical markers, i.e. the grammatical parameter or G parameter, is concerned with morphologically bound and free grammatical morphemes. In a language with a flexible G parameter, one and the same marker can appear in 'clauses' *and* in 'noun phrases', while this is not possible in the case of a rigid system. The combination of the two parameters provides the following four language types:

1. $L_{FLEXIBLE}/G_{FLEXIBLE}$ [L_F/G_F]: Languages with underspecified or precategorial lexemes and with grammatical markers that can occur in 'clauses' and 'noun phrases'
2. $L_{FLEXIBLE}/G_{RIGID}$ [L_F/G_R]: Languages with underspecified or precategorial lexemes and with grammatical markers that occur either in 'clauses' or in 'noun phrases'
3. $L_{RIGID}/G_{FLEXIBLE}$ [L_R/G_F]: Languages with lexemes that are specified for certain syntactic slots and with grammatical markers that can occur in 'clauses' and 'noun phrases'
4. L_{RIGID}/G_{RIGID} [L_R/G_R]: Languages with lexemes that are specified for certain syntactic slots with grammatical markers that occur either in 'clauses' or in 'noun phrases'

The present paper will contribute to the understanding of how these idealized types are linguistically realized by looking at languages which show extreme parameter values. Such a

procedure will not only provide examples of how this typology works in practice it will also reveal some theoretical problems that are associated with it. Since this is only possible if the relevant linguistic phenomena are presented with at least some detail, it is needless to say that only a small number of languages can be discussed. Thus, the paper will be limited to the analysis of four languages, namely Late Archaic Chinese (Sino-Tibetan: Sinitic), Khmer (Austroasiatic: Mon-Khmer: Eastern Mon-Khmer: Khmeric), Tagalog (Austronesian: Western Malayo-Polynesian, Meso-Philippine) and Classical Nahuatl (Uto-Aztecan: Aztecan). Each of these languages covers an important part of the above typology. Late Archaic Chinese is a language which is $L_{FLEXIBLE}$ and cannot be assigned a value of the G parameter, Khmer belongs to type 3 and Classical Nahuatl and Tagalog represent two different realizations of type 4.

- Late Archaic Chinese: It will be argued that Late Archaic Chinese is a precategorial language, i.e., a language whose lexical items are not predetermined for occurring in the syntactic slots of nouns and verbs. From that perspective, Late Archaic Chinese is clearly $L_{FLEXIBLE}$. Since the morphological system of word-class distinction that existed at earlier stages of Chinese (preclassical Chinese between the 11th and the 6th centuries BC) has disappeared to a large extent in Late Archaic Chinese, there is no additional grammatical system that would make it possible to address the G parameter. After the loss of parts-of-speech indicating morphology, the syntactic slots of N and V became the only indicators of word class in Late Archaic Chinese.

- Khmer: This language has a rich inventory of derivational affixal morphology which is characterized by a special type of flexibility. The majority of Khmer affixes are flexible at the global level of the functional range they cover. Thus, one and the same affix can produce a noun or a verb. In combination with individual morphological bases, however, the morphemes are rigid, i.e., the word they produce can only be either a noun or a verb. Thus, the prefix *bvN-* derives a noun from a verb in the case of *tùk* 'put on one side, keep' > *bɔntùk* 'cargo, load, one's work duty', while it produces a causative verb in the case of *co:l* 'enter' > *bɔɲco:l* 'cause to enter, include'. From the perspective of the global function of most of its morphemes, Khmer is a type 3 language that turns into a type 4 language as soon as one looks at its word-class behaviour if a morpheme is combined with an individual base.

  This situation is due to a special historical situation in which the dissyllabic structure of Khmer words provides a rich source for generating new words that is blocked by contact with languages such as Thai, Vietnamese and Chinese which use alternative word-formation processes that are preferred by the speakers of Khmer.

- Classical Nahuatl, Tagalog and extreme rigidity: In terms of Hengeveld (2004: 536), a language that belongs to the type of extreme lexical rigidity "would be a language that has verbs only". This follows directly from the parts-of-speech hierarchy, in which the highest position, i.e. "head of predicate phrase", is the only option that is available if the language has only one word class. From such a perspective, it is thus only natural to conclude that languages belonging to the type of *extreme* rigidity are always verbal. However, a closer look at this type of grammatical rigidity reveals that this is not the whole story. This can be shown from two different perspectives.

  The first perspective is concerned with the morphosyntactic pattern by which the distinction between main predication and argument is marked on the lexemes that occur in the head-of-predicate-phrase slot. There are extreme languages whose elements in that slot follow morphosyntactic patterns that are cross-linguistically associated with verbs, while

others follow the pattern of nominal derivation that is integrated into an equational construction (Bisang 2011). Classical Nahuatl belongs to the former subtype, which is called "omnipredicative" by Launey (1994), Tagalog belongs to the latter subtype (Himmelmann 1987, 2005, 2007). The case of Tagalog shows that even in a system of extreme lexical rigidity, the grammatical morphology (G_{RIGID}) that relates predicates to arguments can have nominal properties.

The second perspective has to do with the extent to which the morphological inventory can be applied to action-denoting and object-denoting lexemes. In Classical Nahuatl as well as in Tagalog, action-denoting lexemes can take the full morphological inventory available for the head-of-predicate-phrase slot, while this is not possible for object-denoting lexemes. Thus, both languages show that at a purely morphological level the noun/verb distinction matters even though it is irrelevant for the mapping of the lexemes carrying that morphology onto syntax. The fact that the distinction between noun and verb is not fully conflated cannot only be shown for languages of type 4, Broschart (1997) presented similar facts for Tongan, a type 2 language in his analysis.

In the remainder of this paper, the individual languages and their type of flexibility will be discussed in separate sections. Section 2 will deal with Late Archaic Chinese in terms of types 1 and 2 and the irrelevance of the G parameter. Section 3 presents Khmer with its morphology that can be both $D_{FLEXIBLE}$ or G_{RIGID}, depending on the level of analysis. Classical Nahuatl and Tagalog will be analysed in section 4 on type 4 languages. Each of these languages will be presented with its parameter values and the theoretical problems that result from its analysis. The paper will end with a short conclusion in section 5 that will argue for a closer integration of the diachronic perspective and processes of morphological change.

2 Late Archaic Chinese: An $L_{FLEXIBLE}$ language whose G parameter cannot be addressed

This section will start with a short account of precategoriality in Late Archaic Chinese in subsection 2.1. Subsection 2.2 will briefly situate this approach in different traditions of dealing with parts of speech in Late Archaic Chinese. Finally, subsection 2.3 will discuss the role of morphology as it is reconstructed for the preclassical period of Old Chinese between the 11th and the 6th centuries BC. It will show how the loss of morphology is responsible for creating a situation in which the G parameter cannot be addressed.

2.1 Precategoriality in Late Archaic Chinese

In Late Archaic Chinese, the language of the classical texts of Confucius, Mencius, Laozi and other authors between the 5th and the 3rd centuries BC, one and the same lexical item can occur in syntactic slots associated with different parts of speech. Thus, the word *měi* 'be beautiful, beauty' takes the slot of a nominal head in (1), while it is found in the V-slot of the transitive argument structure construction in (2):

(1) *měi* is interpreted as a noun (Lunyu 6.14):

宋朝之美

Sòng Zhāo *zhī* *měi*

Zhao.Duke.of.Song ATTR beauty

'the beauty of Zhao, duke of Song'

(2) *měi* is interpreted as a transitive verb (Zuo, Xiang 25):

見棠姜而美之。

jiàn Táng Jiāng ér měi zhī.

see/meet Tang Jiang and beautiful OBJ:3

'He saw Tang Jiang and thought her to be beautiful.'

Starting out from examples like these, it will be argued in this section that Late Archaic Chinese is precategorial, i.e., its lexical items are not preclassified in the lexicon for the assignment to the syntactic slots of nouns and verbs. This will be shown by a look at the transitive and the intransitive argument structure constructions and the principles that determine the mapping of lexical items onto their verbal and nominal slots (for a more detailed account, cf. Bisang 2008a, b, c).

The two argument structure constructions that are relevant for discussing precategoriality in this section are the intransitive argument structure construction (3a) and the transitive argument structure construction (3b) (on intransitive and transitive verbs in Late Archaic Chinese, cf. Pulleyblank 1995: 23-24, 26-28). Intransitive verbs have one NP argument (NP_S) which can precede or follow the verbal slot. Transitive verbs have two argument NPs, an actor (NP_A) and an undergoer (NP_U). If both of their arguments are overt, the actor is in the preverbal slot, the undergoer in the postverbal position. If there is no overt actor argument, an overt undergoer argument can occur in either position depending on constraints of information structure.

(3) Argument structure constructions in Late Archaic Chinese:

a. Intransitive:	(i)	NP_S	V	*or*:
	(ii)		V	NP_S
b. Transitive:		NP_A	V	NP_U

In most languages, there is a clear-cut distinction between verbs that can be used intransitively and those that can be used transitively (cf. the typologically thorough study by Nichols et al. 2004 on transitivizing and detransitivizing languages). If a verb can change its category, this is tied to some morphological changes such as transitivization or detransitivization. In Late Archaic Chinese, almost any verb occurring in the verbal slot of the intransitive construction can also occur in the verbal slot of the transitive construction. In this case, the transitive argument construction contributes causative meaning. The S-role of the intransitive construction becomes the U-role in the transitive construction and the transitive construction adds the A-role. For a more precise description, it is necessary to distinguish between S-arguments with control over the verbal action [+con] and those with no control [-con]. Non-control intransitive predicates will be interpreted in terms of the operators CAUSE and BECOME, while control predicates will only get the CAUSE operator (4a). With non-control predicates, there is a second interpretation, which is called "putative" in sinological literature and produces the meaning of 'consider NP_S to be V' or 'treat NP_S as V' (4b):

(4) Meaning contribution of the transitive argument structure construction
(Notation from Role and Reference Grammar: Van Valin and LaPolla 1997, Van Valin 2005):

a. Causative interpretation:

$\mathbf{V_{intr[-con]}}'(NP_S) \rightarrow NP_A$ [CAUSE [BECOME $\mathbf{V_{intr[-con]}}'(NP_{U(S)})$]]

$\mathbf{V_{intr[+con]}}'(NP_S) \rightarrow NP_A$ [CAUSE $\mathbf{V_{intr[+con]}}'(NP_{U(S)})$]

b. Putative interpretation:

$\mathbf{V_{intr[-con]}}'(NP_S) \rightarrow NP_A$ [CONSIDER/TREAT AS $\mathbf{V_{intr[-con]}}'(NP_{U(S)})$]

In the following two examples, we find the non-control verb *xiǎo* 'be small' in causative (CAUSE plus BECOME) interpretation (5) and in putative (CONSIDER) interpretation (6):

(5) A [-con] verb in causative interpretation (Han Feizi 8.2):
鼻大可小小不可大也。
*Bí dà kě **xiǎo**,xiǎo bù kě **dà** yě.*
nose big can small small NEG can big EQ
'If a nose is big, one can make it smaller, if it is small, one cannot make it bigger.'

(6) A [-con] verb in putative interpretation (Mencius 7 A 24.1):
孔子登東山而小魯，登太山而小天下。
*Kǒng-zǐ dēng Dōng Shān ér **xiǎo** Lǔ, dēng Tài Shān*
Confucius ascend East Mountain and small Lu ascend Tai Shan
*ér **xiǎo** tiānxià.*
and small beneath.the.heavens/the.world
'Confucius ascended the Eastern Mountain and Lu appeared to him small [and he considered Lu to be small], he ascended the Tai Mountain and all beneath the heavens appeared to him small.'

The verbal slot of the transitive/intransitive argument structure construction does not only take action-denoting lexemes and property-denoting lexemes in Late Archaic Chinese, it can equally be filled by object-denoting lexemes. The meaning of this type of utterances can be derived systematically from combining the meaning of the cognitive subcategory to which the lexeme in the verbal slot belongs with the meaning contributed by the construction itself, which basically remains the same as described in (4). In detail, the cognitive subcategories that are relevant to account for the meaning of object-denoting lexemes in the verbal slot are (i) human beings (person-denoting lexemes), (ii) instruments/man-made objects, (iii) sense organs, (iv) places and buildings, (v) pronouns of 1st and 2nd person and (vi) numbers and measures (cf. Bisang 2008b). For the sake of brevity, only subcategory (i) will be illustrated with some examples in this paper (for more details, cf. Bisang 2008b).

The interpretation of person-denoting lexemes (PDL) in the verbal slots of the intransitive and the transitive constructions can be systematically accounted for in terms of the following rules (INT stands for 'used in the intransitive argument structure construction', TR stands for 'used in the transitive argument structure construction'):

(7) N: person/function
INT: a. NP_S behaves like a (true) N, NP_S is a (true) PDL
b. NP_S becomes a (true) PDL
TR c. NP_A CAUSE $NP_{U(S)}$ to V_{intr} (be/behave like a [true] PDL)
d. NP_A CONSIDER $NP_{U(S)}$ to V_{intr} (be/behave like a [true] PDL)

The interpretation of the following very famous example from the Analects (*Lunyu*) of Confucius is based on the intransitive argument structure construction. Both of its slots, the nominal slot for NP_S as well as the verbal slot, take the same lexical items (*jūn* 'prince', *chén* 'minister', *fù* 'father', *zǐ* 'son'). Since the lexical item in the verbal slot follows the rule in (7a), we get the interpretation of 'behave like a (true) prince/minister/ father/son'. The fact that the lexical items in this slot have to be interpreted verbally is further corroborated by their negation by *bù/bú* 'not' in lines 3 and 4:

(8) Intransitive argument structure construction (Mencius 5B.3):
[Context: Duke Jǐng of Qí asked Confucius about good government. Confucius replied:]
君君臣臣父父子子・公曰善哉信如君不君臣不臣父不父子不子雖有粟吾得而食諸。

Jūn ***jūn,*** *chén* ***chén,*** *fù* ***fù,***
N:prince V:behave.like.a.prince N:minister V:minister N:father V:father

zǐ ***zǐ.*** *Gōng yuē:* *shàn* *zāi!* *xìn* *rú* *jūn*
N:son V:son duke say good EXCL believe/indeed if N:prince

bù ***jūn,*** *chén* *bù* ***chén,*** *fù* *bú* ***fù,***
NEG V:prince N:minister NEG V:minister N:father NEG V:father

zǐ *bù* ***zǐ*** *sūi* *yǒu* *sù,* *wú* *dé* *ér* *shí* *zhū?*
N:son NEG V:son even.if have millet I get and eat OBJ:3.Q

'Let the prince behave like a prince, the minister like a minister, the father like a father and the son like a son. The duke said: How true! If, indeed, the prince does not behave like a prince, the minister does not behave like a minister, the father does not behave like a father and the son does not behave like a son, even if I have millet [i.e. food], shall I manage to eat it?'

The following two examples illustrate the use of a person-denoting lexeme in the transitive argument structure construction. In (9), the lexeme *chén* 'minister' is interpreted in terms of (7c) as 'to behave towards NP_U as PDL', i.e., 'to serve NP_U in the function of PDL' (cf. Gassmann 1997: 75). Example (10) illustrates the putative use of the person-denoting lexeme *yǒu* 'friend' with the interpretation of (7d) 'NP_A considers/treats NP_U as a PDL (= friend)':

(9) Person-denoting lexeme in the transitive construction: causative (Zuo, Xiang 22.6):
然則臣王乎。

rán *zé* ***chén*** *wáng* *hū?*
be.so then V:behave.like.a.minister king Q
'Since this is so, will you serve the king as a minister?'

(10) Person-denoting lexeme in the transitive construction: putative (Mencius 5B.3):
吾於顏般也則友之矣

wú *yú* *YànBǎn* *yě,* *zé* ***yǒu*** *zhī* *yǐ.*
I to YanBan be thus V:friend OBJ:3 PF
'What I am to Yan Ban, I treat him/consider him as a friend.'

The precategorial use of lexical items in Late Archaic Chinese implies that they all can occur in boh the verbal and the nominal slots without any differences in marking. Although this seems to be true there are significant differences in the frequency with which individual lexical items actually are attested in one of the two slots. There are lexemes that occur with equal frequency in both slots, while others have strong preferences. The person-denoting lexemes strongly prefer the nominal slot but, as is shown in examples (8) to (10), they occur

in the verbal slot and, if they do so, their meaning can be systematically derived from the rules given in (7). As is shown in Bisang (2008a, b, c), the reason for the differences in the frequency of occurrence of lexical items in the two syntactic slots is due to a pragmatic implicature. While lexemes denoting concrete objects stereotypically imply the occurrence in a nominal slot, lexemes denoting abstract objects are more or less equally open to the nominal and the verbal slot. This correlation is expressed by the following rule, where '>' means 'implies stronger nominal inference than':

(11) CONCRETE OBJECTS > ABSTRACT OBJECTS

This hierarchy can be further refined to the following version of the animacy hierarchy:

(12) 1ST/2ND PERSON > PROPER NAMES > HUMAN > NONHUMAN > ABSTRACTS[1]

The hierarchy in (12) is interpreted in terms of I-implicatures (Inference of stereotype) as defined by Levinson (2000). The higher the position of a lexeme is in hierarchy (12), the more likely is the stereotypical implicature that it belongs to the category of object and will thus be assigned to the nominal slot. As is always the case with pragmatic rules, they can be flouted (Grice 1975). Thus, if a lexical item that takes a high position in hierarchy (12) is used in the verbal slot this produces a flouting effect—and that effect is frequently used for rhetoric purposes in Late Archaic Chinese.

A good example of the rhetoric use of flouting the implicature associated with (12) is (13). In this example, the speaker puts the proper name *Wú wáng* 'King Wu' into the verbal slot.

(13) Proper noun in verbal position (Zuo, Ding 10):
公若曰爾欲吳王我乎。
Gōng Ruò yuē ěr yù ***Wú wáng*** *wǒ hū?*
Gong Ruo say you want Wu king I Q
'Gong Ruo said: "Do you want to deal with me as the King of Wu was dealt with?"'
[King Wu was murdered. → 'Do you want to **kill** me?']

A large proportion of the meaning of *Wú wáng* 'king Wu' in the verbal slot can be derived directly from rule (7d) on person-denoting lexemes in the transitive argument structure construction. The formula 'NP_A CONSIDER $NP_{U(S)}$ to V_{intr} (be/behave like a [true] PDL)' yields the more concrete interpretation of 'you CONSIDER me to be king Wu'. The part of mutual knowledge to derive the concrete meaning of (13) is that king Wu was murdered. With this historical background knowledge, the highly dramatic situational meaning of 'Do you want to kill me?' can easily be inferred. And this is what actually happens to Gong Ruo in the sentence following (13) in a rather dramatic situation of regicide.

[1] There are no 3rd person pronouns for actors in Late Archaic Chinese. Sometimes, we find demonstratives in this function.

2.2 A short sketch on the history of approaches to parts of speech in Late Archaic Chinese

The phenomena of the type described above are known under various terms in the Chinese linguistic tradition (for an excellent survey cf. Zádrapa 2009, who also presents a more in-depth study on parts of speech in Late Archaic Chinese). The most frequently used term is probably *huóyòng* 活用 'word-class transition', which exists since the Song period (960 – 1279 AD) and has been subject to various definitions. Its current use in Chinese linguistics goes back to Chen (1922) who describes it in terms of a derivation from the 'original/basic use' (*běnyòng* 本用) of a lexical item to its 'live or transitional use' *huóyòng* [*huó* 'live', *yòng* 'use']. This definition is unproblematic as long as the basic meaning from which the *huóyòng*-interpretation can be derived is clear but it runs into serious problems in the many instances in which one and the same lexical item is used with about the same frequency as a verb and as a noun.

Another term that is used in Chinese linguistics is *jiānlèicí* 兼類詞 'words that share categories'. This term applies to lexical items whose use in various parts-of-speech functions is part of the lexicon and thus presupposes the determination of parts-of-speech properties in the lexicon. In this approach, the *jiānlèicí* belong to a relatively small subset of lexical items that can take on more than one word-class function. The problem with this view is that it is difficult to draw a clear-cut line between lexical items that clearly belong to only one word class but are attested sometimes in another class and those whose use in different parts-of-speech functions has been lexicalized. While the *huóyòn*g-perspective only works with lexical items that are used frequently in one word-class function and sometimes in another one, the *jiānlèicí* -perspective only works if a lexical item occurs frequently enough in more than one function. Between these two perspectives, there are many unclear cases which cannot be accounted for from either perspective. The precategoriality approach does not depend on frequency of occurrence and offers thus a different pragmatics-based solution to the understanding of parts of speech in Late Archaic Chinese.

Yang and He (1992) presented a statistical study of Late Archaic Chinese texts which show that there is a group of lexical items that is consistently used in one word-class function only. These findings are certainly interesting but they do not prove against precategoriality for at least the following two reasons:

- If an object-denoting lexeme is never attested in the verbal slot this is due to the fact that it is only used in its stereotypical meaning throughout the whole corpus of Late Archaic Chinese texts. This is no argument against its potential of occurring in the verbal slot.
- The semantics of an object denoting lexeme in the verbal slot can be derived very precisely as is shown by (7). In English, which is well-known for its conversion (cf. Clark and Clark 1979), the semantics are much more complex and there is an impressive number of instances in which the correlation between the nominal and the verbal use is purely lexically determined.

The regularity with which the meaning of an object-denoting lexeme in the verbal slot can be derived even with lexemes that are very rarely attested in this slot can be shown from examples such as (14). In this passage from Gongsun Longzi, a Classical Chinese logician, the words *yáng* 'sheep' and *niú* 'ox' are not only used as nouns, they also occur in the V-slot of the transitive argument structure construction followed by the object pronoun of the 3rd

person singular *zhī*. Their meaning in the V-slot follows (7d) NP_A CONSIDER $NP_{U(S)}$ to V_{intr} (be/behave like a [true] PDL:

(14) *niú* 'N: ox/ V: consider as an ox'/ *yáng* 'N: sheep/ V: consider as a sheep' (Gongsun Longzi, cf. Graham 1957: 161):

羊有角牛有角。牛之而羊也羊之而牛也未可

yáng	*yǒu*	*jiǎo,*	***niú***	*yǒu*	*jiǎo,*	***niú***		*zhī*	*ér*
sheep	have	horn	ox	have	horn	V:consider.an.ox		OBJ:3	although

yáng	*yě*	***yáng***		*zhī*	*ér*	***niú***	*yě,*	*wèi*	*kě.*
sheep	EQ	V:consider.a.sheep		OBJ:3	although	ox	EQ	not.yet	possible

'Sheep have horns and oxen have horns. To consider [something] an ox although it is a sheep and to consider [something] a sheep although [it] is an ox is inadmissible.'

Western approaches take different positions to the problem of parts of speech in Late Archaic Chinese. As early as in 1878, von der Gabelentz pointed out that the "vast majority of Chinese words can [...] belong to many different grammatical parts of speech"[2]. Later on in his *Chinese Grammar* (von der Gabelentz 1881: 113), he distinguishes semantics-based categories from syntax-based categories. The semantics-based categories follow ontological criteria similar to those used in Croft's (1991, 2000, 2001) conceptual space for parts of speech (object, property, action) and are referred to by German parts-of-speech terminology ("Hauptwort", "Eigenschaftswort", "Zeitwort"). The syntactic properties of lexical items are defined by their positional potential and are referred to by Latin-based terminology. Thus, lexical items occurring in the verbal slot are "verbs" (and so on for "nouns" and "adjectives"). As von der Gabelentz (1881: 113) points out, "the [semantic] category is an unchangeable property of a word, while the [syntactic] function is subject to change with many words"[3]. Von der Gabelentz's (1881) approach is based on the observation that there are many words whose word-class properties are not rigidly determined in the lexicon. For that reason, he aims at presenting rules that help determining to which syntactic function a word in a particular utterance belongs. His approach is thus similar to the one adopted in this paper.

While von der Gabelentz (1881) clearly looks at lexical items of Late Archaic Chinese from the perspective of parts-of-speech flexibility and tries to provide structural patterns for analysing their function in Chinese texts, most other grammars look at parts of speech in Late Archaic Chinese only briefly and take parts-of-speech determination in the lexicon more or less for granted. A good example is Pulleyblank (1995), who addresses the word-class issue as follows: "In spite of the traces of morphology that can be discerned, words in Classical Chinese are not formally marked for grammatical function. Nevertheless, in their syntactic behaviour they do fall into distinct classes that correspond to such categories as nouns, verbs and adjectives in other languages" (Pulleyblank 1995: 12). Later on, Pulleyblank (1995: 26) states that examples such as (8) and (13) "although not very common, must be regarded as part of the syntactical possibilities of nouns in general" and additionally points out that "particular nouns have acquired special meanings as verbs which must be treated as lexical items" (Pulleyblank 1995: 26). The claim adopted in this paper is that there is a regular

[2] This is my translation. The German original runs as follows: "Die ungeheure Mehrzahl der chinesischen Wörter kann [...] sehr verschiedenen Redetheilen angehören" (von der Gabelentz 1878: 648-649).

[3] This is again my translation. The German original runs as follows: "Die Kategorie ist also dem Worte unwandelbar anhaftend, die Function bei vielen Wörtern wechselnd" (von der Gabelentz 1881: 113).

pattern that accounts for how object-denoting lexemes are interpreted in the verbal slot. The regularity of that pattern plus the observation that the use of object-denoting lexemes in the verbal slot is not that uncommon[4] led me to the conclusion that lexical items in Late Archaic Chinese are precategorial. This does not exclude certain lexicalizations in individual cases as in the following examples quoted by Pulleyblank (1995: 26): *chéng* 'N: wall; V: to wall a city' or *ju□n* 'N: army; V: to encamp'. If precategoriality is a basic property of lexical items and if their assignment to the verbal or nominal slots is governed by pragmatics it is to be expected that certain object-denoting lexemes get specific interpretations if they occur in recurrent situations such as setting up a city wall or the encampment of troops in the above examples. This is the normal development from conversationally-based pragmatics to conventionalized semantics in processes of lexicalization. In addition, there is no reason why more regular interpretations in terms of 'consider X a wall' or 'change X into an army' should not be possible in principle.

2.3 Late Archaic Chinese between type 1 (L_F/G_F) and type 2 (L_F/G_R)

Late Archaic Chinese (5th – 3rd centuries BC) represents the last period of Old Chinese (OC, 11th – 3rd centuries BC). In the preclassical period between the 11th and the 6th centuries BC, Chinese had its own morphology (Karlgren 1920, 1957, Sagart 1999, among others). A look at the affixes reconstructed for that period of time reveals that a remarkable proportion of them are not associated with a single slot for parts of speech. Word-class sensitivity only comes in at the level of individual words, i.e. at the level of the lexicon. Thus, if the prefix **s*- marks an object-denoting lexeme such as OC *^{b}*tu?* 'a broom' it changes that base into the verb OC *^{a}***s***-*tu?* 'to broom' (15a). If the same suffix is used in an action-denoting context as in *^{b}*m-lak-s* 'to shoot with a bow' it produces the noun *^{b}***s***-*lak-s* 'open hall for archery exercise' (15b). From a more general point of view, affixation seems to conventionalize the word class of a lexeme, while the occurrence of morphologically unmarked words in a nominal or verbal slot of the argument structure construction is based on conversational implicatures which can be flouted according to the scale of likelihood reflected by the animacy hierarchy as presented in (12). In my view, this difference between the levels of morphology and syntax is due to the fact that the unrestricted application of conversational implicatures at both levels would be communicatively unfavourable because such a constellation increases the number of potential interpretations to an extent which seriously endangers the adequate on-time understanding of an utterance by the hearer.

As will be seen in section 3, the situation in Old Chinese is similar to morphology in modern Khmer whose affixes are flexible on a global level but rigid if they are applied to an individual lexical item. It is thus a type 3 language at the global level and a type 4 language at the level of its impact on individual bases. This will be illustrated briefly by the example of the prefix ***s***-, which derives verbs from object-denoting lexemes (15a) as well as nouns from action-denoting lexemes (15b) (cf. Bisang 2008a 583-585 for this and another example with the suffix **-s*). In addition, the prefix **s*- also marks causativity (15c), directives (15d) and maybe[5] inchoatives (15e; cf. Sagart 1999: 72). Directives are defined in Mei's (1989) paper on the prefix **s*- as "acts or states directed towards external conditions or another person" (quoted from Sagart 1999: 71).

---

[4] On the problem of action-denoting lexemes in the nominal position to satisfy Evans and Osada's (2005) condition of bidirectionality, cf. Bisang (2009: 579-582).

[5] Sagart (1999) uses a question mark with regard to the inchoative function.

(15) a. verbs derived out of object-denoting lexemes (Sagart 1999: 71):

麗	*lì*	MC **lejH*	OC ***ra*re(ʔ)-s*	'a pair, a couple'
灑	*sǎ*	MC **sreaiX*	OC ***a*s-re(ʔ)*	'to divide, bifurcate'
帚	*zhǒu*	MC **tsyuwX*	OC ***b*tuʔ*	'a broom'
掃	*sǎo*	MC **sawX*	OC ***a*s-tuʔ*	'to broom'

b. nouns derived out of verbs (Sagart 1999: 73):

射	*shè*	MC **zyæH*	OC ***b*m-lak-s*	'to shoot with bow'
榭	*xiè*	MC* *zjæH*	OC ***b*s-lak-s*	'open hall for archery exercises'
抴	*yì*	MC **yet/yejH*	OC ***b*lat(-s)*	'to pull'
韘	*xiè*	MC **sjet*	OC ***b*s-hlat*	'leading-string'

c. causatives (Sagart 1999: 70):

順	*shùn*	MC **zywinH*	OC ***b*m-lun-s*	'pliant, obedient'
馴	*xún*	MC **zwin*	OC ***b*s-lun*	'to make (a horse) obedient'
食	*shí*	MC **zyik*	OC ***b*m-lɨk*	'to eat'
飼	*sì*	MC **ziH*	OC ***b*s-lɨk-s*	'to feed (tr.)'

d. directives (Sagart 1999: 71):

易	*yì*	*MC *yek*	OC ***b*lek*	'to exchange'
賜	*cì*	MC **sjeH*	OC ***b*s-hlek-s*	'to give'

e. inchoatives (Sagart 1999: 72):

悟	*wù*	MC **nguH*	OC ***a*ŋa-s*	'to be awake, aware'
蘇	*sū*	MC **su*	OC ***a*s-ŋa*	'to come back to life; to wake up'

Even though it is hard to assess the exact status of word classes in the preclassical period the existence of affixes that assign individual lexical items to particular parts of speech clearly points into the direction of G_{RIGID}. Since it is more difficult to say whether preclassical Chinese was $L_{FLEXIBLE}$ or L_{RIGID}, it may be a type 2 or a type 4 language. Late Archaic Chinese is characterized by the loss of a large part of its former morphology. For that reason, the G parameter is difficult to assess even though it seems plausible that there still existed some residual morphology with the property of G_{RIGID} at that time. Given the above analysis of Late Archaic Chinese in terms of precategoriality, it is an $L_{FLEXIBLE}$ language whose G parameter is irrelevant except for a potential residual set of lexemes in which the occurrence in 'clauses' and in 'noun phrases' is still morphologically expressed. Thus, Late Archaic Chinese is not a clear instance of a type 1 language but it is neither a type 2 language. This particular status is the result of a diachronic morphophonological process which led to the reduction of morphology (on the general consequences of this process for the typology of later stages of Chinese, cf. Xu 2006). As a consequence, word order and the distinction between the nominal

and the verbal slots became the only possibility to determine the word class to which a lexical item belongs.

3 The G parameter and morphology in Khmer

Khmer is one of the languages of East and mainland Southeast Asia with a relatively rich derivational morphology that mainly consists of prefixes and infixes (Jenner and Pou 1980/1981, Bisang 1992: 447–472). According to some authors (Lewitz 1967: 121; Haiman 1998), Khmer morphology is productive at least in the case of some morphemes. And in fact some affixes seem to have reached a certain degree of productivity that is evidenced by the fact that they can be found incidentally in new combinations not attested in Khmer dictionaries. Probably the most likely candidate for a productive affix is the infix *-vmn-/vN-* (cf. (17) below). In general, however, morphology is not productive. Moreover, there is a tendency in spoken Khmer to use simple words with no affixation (Michel Ferlus, p.c.). Morphology is thus mainly a phenomenon of the written language.

The richness of Khmer morphology is in strong contrast to the relatively few functions it can express. The 28 affixes discussed by Jenner and Pou (1980/1981; summary in Bisang 1992: 457) express only three main functions. The first two of them are nominalization and causativization/transitivization. The third function is specialization, a term created by Jenner and Pou (1980/1981) that covers all instances in which the meaning of the affixed form is lexically more specific than the meaning of the morphological base. Since one and the same affix can have more than one of these functions, the majority of Khmer affixes lack parts-of-speech sensitivity in the sense that the same morpheme derives a noun with one lexical base and a verb with another base. Thus, the polyfunctionality of Khmer derivational affixes leads to parts-of-speech flexibility across morphological bases but they do not generate ambiguities with individual bases. The list in (16) presents the most frequent affixes that are not lexically specified for the use in nominal or verbal syntax. This can be seen from the parts-of-speech indications in the rightmost column[6]. In addition to the three main functions, the markers in this list express two more form-specific functions—one of them is affectedness[7] (*k-*), the other is reciprocity (*prə-*).

(16)

	Affix/Function	Base	Base + Affix	Word class of Base + Affix
k-	Nominalization:	*baŋ* ‘use sth, shade/cover sth’	*kbaŋ* ‘visor, fender, guard’	**noun**
	Affectedness:	*ca:y* ‘scatter, spend’	*khca:y* ‘scattered, spilt’	**verb**
p-	Nominalization:	*lèːŋ* ‘play’	*phlèːŋ* ‘music’	**noun**
	Causativization/ Transitivization:	*dac* ‘break [intr.], be torn apart’	*phdac* ‘break [tr.]’	**verb**
	Specialization:	*haəm* ‘swollen’	*phaəm* ‘swollen, be pregnant’	**verb**
s-	Nominalization:	*pìːən* ‘pass over, traverse’	*spìːən* ‘bridge’	**noun**
	Specialization:	*rù:t* ‘hurry’	*srò:t* ‘swiftly and directly’	**verb**

[6] Since Khmer is an adjectval-verb language in terms of Schachter (1985), property-denoting words are called verbs in (16).

[7] Schiller (1985: 84) describes this function as follows: “By [affected] I mean that the predicating element bearing this feature conveys to its logical argument a sense that the predication has already taken place and that the argument has been affected by the process referred to by the predicate”. This function is also expressed by the prefix *L-*.

		dɔh ‘free [verb]’	*sdɔh* ‘spit, eject’	**verb**
prə-	Nominalization:	*cheh* ‘catch fire’	*prəcheh* ‘wick’	**noun**
	Causativization/ Transitivization:	*kaət* ‘be born’	*prəkaət* ‘cause, produce’	**verb**
	Reciprocity:	*kham* ‘bite’	*pbɔntùk* ‘cargo, load’	**noun**
	Causativization/ Transitivization:	*baek* ‘break [intr.]’	*bɔmbaek* ‘break [tr.]’	**verb**
	Specialization:	*lèːŋ* ‘play’	*bɔŋlèːŋ* ‘entertain’ [by playing music]	**verb**
kvN-	Nominalization:	*cah* ‘old’	*kɔɲcah* ‘old person’ [derogatory]	**noun**
	Specialization:	*bak* ‘broken’ [in general]	*kɔmbak* ‘broken’ [body part]	**verb**
svN-	Nominalization:	*bɔ̀ːk* ‘peel, strip off’	*sɔmbɔ̀ːk* ‘shell, husk, skin’	**noun**
	Specialization:	*cèːk* ‘separate, divide’	*sɔɲcèːk* ‘be open, yawn’	**verb**
-vmn-	Nominalization:[8]	*dam* ‘plant [v.]’	*dɔmnam* ‘plant [n.]’	**noun**
vN-	Causativization/ Transitivization:	*sʔaːt* ‘be clean’	*sɔmʔaːt* ‘to clean’	**verb**
	Specialization:	*krùː* ‘teacher’	*kùmruː* ‘model, example’	**noun**

The most frequent affixes that can produce nominal and verbal forms in (16) are *-vmn-*/ *-vN-*, *prə-*, *bvN-*. Each of them will be described in some more detail to provide more concrete evidence of their flexibility. After these examples, it will be shown that there are some exceptions in which individual affixes show parts-of-speech sensitivity. At the end of this section it will be briefly outlined how this particular situation developed from a combination of phonological properties of the word plus contact with other languages that had much less or no morphology.

As was pointed out above, the infix *-vmn-/-vN-* is the most likely candidate for productivity. However, no matter how productive it is, it produces a noun with some bases (example (17a) plus some instances of (17c)) and a verb with other bases (example (17b) plus some instances of (17c)):

(17) Functions of the syllabic infix *-vmn-/-vN-*:

a. Nominalisation:

Results:	*khɔp* ‘to pack’	→	*kɔɲcɔp* ‘parcel [n.]’
Instruments:	*tbaːɲ* ‘weave [v.]’	→	*dɔmbaːɲ* ‘equipment for weaving’
Abstracts:	*thŋùən* ‘be heavy’	→	*tùmŋùən* ‘heaviness, weight’
	lʔɔː ‘be beautiful’	→	*lùmʔɔː* ‘beauty; embellishment’
Agentives:	*taːŋ* ‘go in place of’	→	*dɔmnaːŋ* ‘representative’

b. Causativization/Transitivization:
sʔaːt ‘be clean(ed)’ → *sɔmʔaːt* ‘to clean’

[8] The form *-vmn-* occurs with basic syllables of the form CV(C), whereas *-vN-* is used with $C_1C_2V(C)$. With C_1C_2VC-bases, the infix occurs between C_1 and C_2 (cf. Huffman 1967).

	thlɛ̀ək ‘fall’	→	*tùmlɛ̀ək* ‘cause to fall, put down’
	slap ‘die’	→	*sɔmlap* ‘to kill’
	rədɔh ‘be freed, free from’	→	*rùmdɔh* ‘to free [tr.]’

c. Specialization:

krù: ‘teacher’	→	*kùmrù:* ‘model, example’
lʔɔ:ŋ ‘dust’	→	*lùmʔɔ:ŋ* ‘fine dust, pollen’
kɔt ‘note down’	→	*kɔmnɔt* ‘fix [a date] [v.], record [n./v.], inscription [n.]’
chlɔ:ŋ ‘to cross’	→	*cɔmlɔ:ŋ* ‘to copy’
tè: ‘not (clause-final negation)’	→	*tùmnè:* ‘be empty, be free, have time’

The functions expressed by the infix *-vmn-/-vN-* are partly determined by the ontological status of the base morpheme. If the base denotes an object only the function of specialization is possible. With bases denoting actions or properties, the infix can either produce a noun or a verb or even a few lexically unspecified forms (cf. *kɔmnɔt* in (17c)). As is typical of several nominalizers in Khmer (also cf. the functions of the infix *-b-* in (21)), the infix *-vmn-/-vN-* covers a wide range of different semantics such as results, instruments, abstracts and agentives.

The prefix *prə-* has three basic functions. It marks causativization/transitivization (18a), change of word class (18b) and reciprocity (18c). The only case where some prediction is possible is reciprocity because this function is only possible with action-denoting bases and its result is always a verb. The other two functions do not allow any predictions. Thus, the semantic/cognitive status of the base is not related to the parts-of-speech properties of the derived word. In addition, the function of change of word class (18b) is possible in both directions, i.e. from noun to verb or from verb to noun—a very clear-cut indication of the flexibility of that affix.

(18) Functions of the prefix *prə-*:

a. Causativization/Transitivization:

kaət ‘be born’	→	*prəkaət* ‘cause, produce’
do:c ‘be like’	→	*prədo:c* ‘compare’
mù:l ‘be round’	→	*prəmo:l* ‘gather together’
cṳ̀m ‘a round, a turn around’	→	*prəcṳ̀m* ‘assemble [tr. and intr.]’

b. Change of word class:

cheh ‘catch fire, be on fire’	→	*prəcheh* ‘wick’
vɛ̀:ŋ ‘be long’	→	*prəvaeŋ* ‘length’
yùt̀(ìh) ‘combat [n.]’	→	*prəyot̀(ìh)* ‘to fight, attack’

c. Reciprocity:

kham ‘to bite’	→	*prəkham* ‘to bite each other’
chlùəh ‘to quarrel’	→	*prəchlùəh* ‘to squabble together’
deŋ ‘persue, chase away’	→	*prədeŋ* ‘to persue each other, to compete’

The prefix *bvN-* marks the three functions of causativization/transitivization (19a), nominalization (19b) and specialization (19c). It does not strictly specify word class, since its products can either be nouns or verbs, but its use seems limited to a large extent to bases denoting actions or properties:

(19) Functions of the prefix *bvN-*:

a. Causativization/Transitivization:

baek 'break [intr.]'	→	*bɔmbaek* 'break [tr.]'
rìən 'learn'	→	*bɔŋrìən* 'teach'
coːl 'enter'	→	*bɔŋcoːl* 'cause to enter, include'
chùː 'be ill'	→	*bɔŋchùː* 'hurt (mentally)'
phlëc 'forget'	→	*bɔmphlëc* 'make forget, forget intentionally'

b. Nominalization:

tùk 'put on one side, keep'	→	*bɔntùk* 'cargo, load, one's work duty'
vèc 'to parcel up'	→	*bɔŋvèc* 'package [n.]'

c. Specialization:

lèːŋ 'to play'	→	*bɔnlaeŋ* 'to play, amuse'
kac 'to break [tr.]'	→	*bɔŋkac* 'to falsely accuse'

Even though the lack of parts-of-speech distinctions holds for a large number of Khmer affixes there are also some exceptions. Thus, all instances marked by the prefix *crə-* are verbs:[9]

(20) Words with the pefix *crə-*:

lòːt 'jump'	→	*crəlòːt* 'jump up'
mùc̀ 'sink, immerse oneself'	→	*crəmùc̀* 'sink [tr.]'
lɤ̀ːh 'cross [bridge, border, etc.]'	→	*crəlaəh* 'go beyond the law, lawless'
mùːl 'be round'	→	*crəmòːl* 'tangled, intertwined'
bac 'bundle, fagot'	→	*crəbac* 'squeeze with the fingers, massage'

While affixes that produce derivations of exclusively verbal function are rare, there are no less than five affixes which exclusively or almost exclusively form nouns (*m-*, *N-*, *-b-*, *-m-*, *-n-*). To conclude this section, the infix *-b-* will be briefly described. Lewitz (1976) lists 72 words marked by this infix and states that 70 of them are nouns derived from verbs. However, this statement underestimates the number of words marked by *-b-* that can also be used as verbs at least in a secondary function. Example (21) presents the main functions of *-b-* in its nominalizing function. As can be seen, these functions largely overlap with the nominalizing function marked by the infix *-vmn-*/*-vN-* in (17). This is another good indicator of the lexical character of Khmer morphology and its limited degree of productivity.

[9] This may also be accidental because the prefix *crə-* is not productive. In fact, the list in (20) may cover all the instances of its occurence.

(21) The nominalizing functions of the infix *-b-* (Lewitz 1976):

Process:	*yùm* 'to cry'	→	*yò:bam* 'cry [n.]'
Result:	*rɔ̀əm* 'to dance'	→	*rəbam* 'dance [n.]'
Abstract:	*lùən* 'quick'	→	*lbùən* 'speed'
Agentive:	*rì:əl* 'spread, expand'	→	*rəba:l* 'guard, watchman; also used as V'
Instrument:	*rɛ̀əŋ* 'bar the way'	→	*rəbaŋ* 'barrier'
Specialization:	*rò:y* 'fall [petals of dying flowers]'	→	*rəbaoy* 'become sparse [about fruits towards the end of season]'

This last example shows that there is parts-of-speech related morphology in Khmer. The problem is that this is not a consistent property of the language. Only 6 of the 28 affixes discussed in Jenner and Pou (1980/1981) are rigid, the other 22 are flexible. Moreover, the lack of parts-of-speech sensitivity is characteristic of relatively frequent affixes such as *-vmN-*/*-vN-* (17), *prə-* (18) and *bvN-* (19). Thus, there is good evidence for the flexibility of Khmer affixes and against their general association with individual word classes. The factors that led to such a situation will be discussed in the remainder of this section (cf. Bisang 2001: 195 – 200).

For a better understanding of the flexibility of Khmer affixes it is neceesary to look at the phonological properties of words. Since Henderson (1952) Khmer is well-known for its canonical iambic or sesquisyllabic word structure. Thus, a word consists either of one syllable with the structure (22) or of two syllables with a reduced first syllable as in (23). Henderson (1952) uses the terms major syllable and minor syllable for the two syllable types.

(22) Structure of major syllables
C(C)VC or C(C)VV(C)[10]

(23) Structure of minor syllables (Huffman 1967: 45)
C_1 - (C_2) - V - (C_3), C_1: every possible C, but only *p*, *t*, *c*, *k*, *s* in front of C_2.
C_2: only *r*
V: every possible short vowel, but only *ə*, *a*, *o*, *ɔ* after C_2.
C_3: every C except *b*, *d*, *r*.

From the point of view of morphology, not all the minor syllables in (23) are of equal importance. The bisyllabic word structures which are relevant for morphology can have the following structures (also cf. Nacaskul 1978):

(24) a. Cə-major syllable
b. Crə-major syllable
c. CvN-major syllable, C: any consonant
v: *ɔ, ùə, ù*
N: *m, n, ɲ, ŋ* (homorganic pronunciation according to C_1 of the major syllable)

[10] Huffman (1967) also mentions two loan words with the structures CCCVC and CCCVVC: *stha:n* 'place' and *lkhaon* 'theatre'.

The minor syllables in structure (24) are reduced in spoken language according to the following cline (Huffman 1967: 48-49), in which the pronunciation given in the first column is very careful and formal, whereas the pronunciation in the last column is colloquial and very informal:

(25) a.	*rəbɔːŋ*	*rəbɔːŋ*	*ləbɔːŋ*	'fence'
b.	*prɔtɛ̀əh*	*prətɛ̀əh*	*pətɛ̀əh*	'meet'
c.	*kɔndaːl*	*kəndaːl*	*kədaːl*	'middle, center'

The process of reduction presented in (25) is very common in colloquial speech. Its consequence is that the distinction between bisyllabic words with a minor syllable and monosyllabic words with two onset consonants becomes obsolete. Thus, (25b) *ptɛ̀əh* 'meet' cannot be distinguished phonetically from *phətɛ̀əh* 'house' with a CC onset. The convergence of bisyllabic words with minor initial syllable and monosyllabic words with C_1C_2 onset creates a situation in which almost any consonant in the CC onset may be analysed as a prefix derived from a bisyllabic word with a prefix on *Crə-* or *CvN-*. This creates an enormous potential of reanalysis that contributes to the generation of new morphemes in Khmer which is directly linked to the sesquisyllabic structure of its words (on this topic, cf. Haiman 1998). Once an individual consonant C_x within a C_xC onset has been analysed as a prefix, it can then also be extended to $C_xrə$- or C_xvN- by analogy to *Crə-* or *CvN-* in general.

If the continuum between minor syllables of the type *Crə-* or *CvN-* and CC onsets provides such a high potential of morphological productivity one may wonder why Khmer never developed a fully productive and functionally more consistent morphological system, i.e., a system in which one particular affix marks one particular function. The reason seems to be language contact (on the several hundred years of intensive language contact between Thai and Khmer, cf. Huffman 1973, and more recently Enfield 2003). The morphological potential of Khmer was not used because alternative patterns were adopted from other Southeast Asian languages. This can be illustrated by the patterns employed for the formation of agent nouns. In principle, there is derivational morphology such as the infix *-m-* which can be used for producing agentive nouns such as *smòːm* 'beggar' from *sòːm* 'ask, ask a favour', *chmam* 'guard, n.' from *cam* 'wait for, guard, keep' and *chmùːəŋ* 'business-man' from *cùːəŋ* 'do business'. In contemporary Khmer, however, this pattern is replaced by non-morphological alternatives that are used in Thai, Vietnamese, Chinese and many other languages of the area. Thus, agentive nouns are productively formed by the noun *nɛ̀ək* 'person' in the head position as in *nɛ̀ək-daə(r)* [person-walk] 'pedestrian', *nɛ̀ək-taeŋ* [person-compose/write] 'author, composer, writer' and *nɛ̀ək-chlɔ̀ːp* [person-go stealthily to watch someone] 'spy, snoop'.

If one takes the potential provided by the phonological properties of the Khmer word and the influence of language contact together the flexibility of Khmer derivational affixes seems to be due to the fact that the alternative syntax-oriented pattern of derivation overrode the morphological pattern at a time when it had not yet developed into a more consistent system in which individual affixes were rigidly associated with a single syntactic slot either for nouns or for verbs.

4 Parts-of-speech distinctions in L_{RIGID}/G_{RIGID} languages — the case of Classical Nahuatl and Tagalog

4.1 Classical Nahuatl as an omnipredicative language

Classical Nahuatl is a language with a lexically extremely rigid parts-of-speech system and with a grammatically rigid system that patterns cross-linguistically with verbal morphology. Thus, the lexemes occurring in the head-of-predicate-phrase slot show agreement with the single argument of intransitive verbs (S)/actor-argument of transitive verbs (A) vs. undergoer-argument of transitive verbs (U). This is the type of language Sasse (1993a, b) had in mind when he described Iroquoian in general and Cayuga in particular. Even though it turned out later that his analysis of Iroquoian was untenable (Mithun 2000) this does not exclude the existence of such a type of language elsewhere. As was shown by Launey (1994), Classical Nahuatl is such a language. Since it only uses predicative structures, he called it an "omnipredicative" language. In addition, Launey (1994) also convincingly argues that omni-predicativity does not necessarily exclude parts-of-speech distinctions.

The first half of this section will present the lexical rigidity of Classical Nahuatl and the verbal agreement morphology by which the lexemes in the head-of-predicate-phrase slot are marked. The second part will illustrate the relevance of noun/verb distinction as it is reflected in the extent to which the morphological inventory can be applied to action-denoting and to object-denoting lexemes.

In Classical Nahuatl, the person markers for intransitive arguments (S) and for active arguments (A) of transitive predicates are identical to the person markers for predicative constructions with lexemes that denote physical objects. Thus, the prefixes *ni-* (1SG), *ti-* (2SG) and *ø-* (3SG) can be equally prefixed to action denoting lexemes (*ni-chōca* 'I cry', *ti-chōca* 'you cry', *ø-chōca* 's/he cries') and to lexemes denoting physical objects (*ni-tīcitl* 'I am a doctor', *ti-tīcitl* 'you are a doctor', *ø-tīcitl* 's/he is a doctor'). If this analysis is correct a seemingly simple sentence like (26) consists of two predications: 's/he cries' and 'it is a child':

(26) *ø-chōca* *in* *ø-piltōntli.*
3-SG-cry LNK 3SG-child
'S/He cries — it is a child'

The analysis in (26) shows that even the lexeme denoting the physical object 'child' is not a simple noun but a predicate and thus supports the omnipredicative interpretation of Classical Nahuatl (on the analysis of the linker *in*, cf. below). Further evidence for that analysis will be given after some more information on morphology. As can be seen from table 1, action-denoting lexemes have another set of person markers for undergoers (U) in addition to the person markers for S/A, while lexemes denoting physical objects (= object-denoting lexemes) additionally have possessive affixes:

	Action-denoting lexemes			Object-denoting lexemes	
	S/A	U	Reflexive	S	Possessive
1SG	*t(i)-*	*-nēch-*	*-n(o)-*	*n(i)-*	*-n(o)-*
2SG	*t(i)/x(i)-*	*-mitz-*	*-m(o)-*	*t(i)-*	*-m(o)-*
3SG	*ø-*	*-c-/-qu(i)-*	*-m(o)-*	*ø-*	*-ī-*

1PL	*t(i)-*	*-tēch-*	*-t(o)-*	*t(i)-*	*-t(o)-*
2PL	*aM-/x(i)-*	*-amēch-*	*-m(o)-*	*aM-*	*-am(o)-*
3PL	*ø-*	*-quiM-*	*-m(o)-*	*ø-*	*-īM*

Table 1: Person agreement markers on action-denoting lexemes and on object-denoting lexemes (adapted from Launey 1994: 10–11)

An utterance like 'S/He eats it' thus takes the form of *ø-qui-cua* [3SG:A-3SG:U-eat] as in the following example:

(27) Launey (1994: 37):

ø-qui-cua	*in*	*piltōntli*	*in*	*nacatl.*
3SG:A-3SG:U-eat	LNK	child	LNK	meat

'The child eats the meat.'

In an omnipredicative analysis, (27) consists of the three predications 'S/He eats it', 'It's a child' and 'It's meat'. The object-denoting lexemes are thus again predicates of their own and the semantic relation between them is determined by coindexation between their person markers and the person markers of the action-denoting lexeme:

(27') $\underbrace{ø_i\text{-}qui_j\text{-}cua}$ *in* $\underbrace{ø_i\text{-}piltō\bar{n}tli}$ *in* $\underbrace{ø_j\text{-}nacatl.}$

'S/He eats it' 'It is a child' 'It is meat.'

The overall omnipredicative analysis of examples such as (26) and (27) is further corroborated by various other facts. One of them is word order. The positions of the individual predications of object-denoting and action-denoting lexemes are governed purely by information structure. Thus, any possible order of S (subject/actor-argument), O (object/undergoer argument) and V (verb/main predicate) is possible even if the most frequently attested word-order sequences are VSO and SVO (Launey 1994: 41). The topic (*thème* in terms of Launey 1994: 134-136) takes the clause-initial position and is marked by the linker *in*. Sequences with two arguments in front of the predicate (SOV, OVS) are rare because sentences with two topics are rare in general. Finally, the optionality of arguments is another consequence of the discourse-dependence of sentence structure in Classical Nahuatl. Arguments can always be omitted if they are known from context.

A second support of the omnipredicative analysis of Classical Nahuatl is related to person marking. Since the third person is *ø*-marked, the predicative status of both object-denoting and action-denoting lexemes can only be inferred from the overall morphological paradigm of predicative person marking in examples (26) and (27). Thus, there is no overt indication that there really are two predications of equal status. This is different with first and second person marking, since there are examples in which the action-denoting predicate and the object-denoting predicate carry both the same person marker. In (28), this is the marker of the second person plural:

(28) Classical Nahuatl (Launey 1994: 72):

*N-**amēch**-tzàtzilia*	*in*	***an**-tlamacaz-quê.*
1SG:A-2PL:U-implore	LNK	2PL:S-high.priest:PL

'I implore you – who are high priests!'

A last corroborative argument is the fact that a word like *piltōntli* 'S/He is a child' cannot be used to call a person. There are special vocative forms for that purpose.

Another factor that needs to be considered in the onmipredicative analysis of Classical Nahuatl is the linker *ni* (LNK), which is analysed by Launey (1994: 122–132) as a 'pivot' that links two predications. Since it is generally used as a demonstrative or as a relative marker, these functions can be used simultaneously on one instantiation of *in* in predicate linking if it is the demonstrative argument of one predication and the relative marker of the other predication. In example (26), *in* is thus a demonstrative in the argument position of 'cry' (*ø-chōca in* 'That one/he/she cries') and the relative marker of the second predicate 'be a child' (*in piltōntli* 'the one who is a child'). A literal translation of (26) may thus look as follows: 'That one cries, the one who is a child'.

The above sketch provides good evidence for the overall omnipredicative structure of Classical Nahuatl and thus for the lexical rigidity and the grammatical rigidity of the parts-of-speech system of Classical Nahuatl. In spite of this, action-denoting lexemes (including property-denoting lexemes) and object-denoting lexemes do not make the same use of the morphology that is available for lexemes in the head-of-predicate-phrase slot. For that reason, there is a noun/verb distinction in Classical Nahuatl. In general, object-denoting lexemes can take only a limited set of morphological markers that can occur in that position, while action-denoting lexemes can take the full range.

The two most prominent properties that provide evidence for the existence of a noun/verb distinction are tense-aspect marking and the number of arguments (Launey 1994: 201). There is a number of tense-aspect markers which can only occur with action-denoting (and property-denoting) lexemes. Some of these markers are *-z* for future (in *chōca-**z*** 'S/He will cry'), *-c* for perfective (in *chōca-**c*** 'S/He cried'), *ō* for completed action (in ***ō*** *chōca-c* 'S/He has cried'), *-ya* for imperfective (in *chōca-**ya*** 'S/He was crying') and *-zquiya* for irrealis (in *chōca-**zquiya*** 'S/He was about to cry' / 'S/He almost cried') (Launey 1994: 28). In addition to the tense-aspect markers, there are modality particles, which can occur with action/property-denoting lexemes as well as with object denoting lexemes. The following example illustrates an object-denoting lexeme (*ø-tīcitl* 'S/He is a doctor') and an action-denoting lexeme (*ø-chōca* 'S/He cries) with the modality particles *ca* (factivity of a state of affairs), *cuix* (question particle), *mach* (evidentiality) and *zan* (exclusivity in the sense of 'only'):

(29) Modality particles in predications with object-denoting lexemes (Launey 1994: 31, 51):

a. *Ca tīcitl.* 'S/He really is a doctor.' / *Ca chōca* 'It's a fact that s/he cries.'

b. *Cuix tīcitl.* 'Is s/he a doctor?' / *Cuix chōca* 'Does s/he cry?'

c. *Mach tīcitl.* 'S/He must be a doctor [inference].' /
Mach chōca 'It looks as if s/he is crying.'

d. *Zan tīcitl.* 'This is only a doctor.' / *Zan chōca* 'S/He does nothing but cry.'

The second property that generates a distinction of nouns and verbs is related to argument structure. As can be seen from table 1, object-denoting lexemes (= nouns) differ with regard to their relational structure from action-denoting lexemes. Object-denoting lexemes have only one relation (S-argument) plus a possessor argument (last column of table 1), while lexemes denoting transitive actions have an additional secondary argument (U in table 1).

Finally, there is a considerable number of morphemes which are clearly associated with nouns. Some of them only occur with object-denoting lexemes, others are used for the

derivation of nouns from action-denoting morphological bases. The most prominent marker which occurs exclusively with object-denoting lexemes is *-tl*/*-tli* [*-ɬ*/*-ɬi*]. It occurs with lexemes denoting humans, animals, plants, natural objects, artefacts, abstracts and processes (*cihuātl* 'be a woman', *oquichtli* 'be a man', *coyōtl* 'be a coyote', *tepētl* 'be a mountain', *oquichyōtl* 'it is manhood, virility', *tlamatiliztli* 'it is science', etc.; cf. Launey 1994: 207–208). Words with that affix can safely be assigned to the class of nouns (there are other classes of object-denoting lexems which take another suffix *-in* or no marker at all). Other indicators of nominality are markers of diminutives (*-tōn-*, cf. *pil-tōn-tli* 'child' in (26) and (27)), augmentatives (*-pōl-*), honorifics (*-cin-*) and derogatives (*-sol-* for objects, *-pōl-* for humans; cf. Launey 1994: 202–203). In addition, there are derivational affixes that can be used to change an action-denoting lexeme into a noun. The most productive form is *tla-...-l-li* as in *tla-cua-l-li* 'food, nourishment' (from the root *-cua-* 'eat') or *tla-'tō-l-li* 'the things said: word, speech, language' (from the root *ìtoa* 'say') (Launey 1994: 266).

4.2 Tagalog—a type 4 language with a noun-based predicative morphology

The extremely rigid grammatical parts-of-speech system of Tagalog patterns cross-linguistically with nominal morphology. For understanding this system, it is necessary to look at the structure of simple declarative sentences and the way in which the relation between the predicate and nominal participants is expressed. For that purpose, the following example with the action-denoting lexeme *bili* 'buy' in the clause initial position will be discussed in some detail:

(30) Tagalog (Foley and Van Valin 1984: 135):

a. *B-um-ili* ***ang*** ***lalaki*** *ng* *isda* *sa* *tindahan*.
/PFV/AT/-buy T man U/A fish LOC store
'*The man* bought fish at the store.'

b. *B-in-ili-ø* *ng* *lalaki* ***ang*** ***isda*** *sa* *tindahan*.
/PFV/-buy-UT U/A man T fish LOC store
'The man bought *the fish* at the store.'

c. *B-in-il-han* *ng* *lalaki* *ng* *isda* ***ang*** ***tindahan***.
/ PFV/-buy-LT U/A man U/A fish T store
'The man bought fish *at the store*.'

d. *Ip-in-ang-bili* *ng* *lalaki* *ng* *isda* ***ang*** ***pera***.
IT/PFV/-buy U/A man U/A fish T money
'The man bought fish *with the money*.'

e. *I-b-in-ili* *ng* *lalaki* *ng* *isda* ***ang*** ***bata***.
BT-/PFV/-buy U/A man U/A fish T child
'The man bought fish *for the child*.'

The action-denoting lexeme in the clause-initial position is marked for aspect (perfective, imperfective and contemplated; cf. Schachter and Otanes 1972: 66), kind of action (cf. Ramos 1974a: indicative, distributive, aptative, social, causative) and for the semantic role of the structure marked by *ang* (with common nouns) or *si* (with proper names). Apart from *ang*, noun phrases, which follow the verb in more or less free word order, can be marked by the

case markers *ng* (*ni* with proper names) and *sa* (*kay* with proper names). The case marker *ng* occurs with actors, undergoers and a subset of instrumentals which are not in the *ang* phrase and thus are not marked for semantic role on the verb. The same marker is also used in possessive constructions.[11] The case marker *sa* is associated with the semantic roles of goal, recipient and location. It is used if a noun phrase in one of these roles is not marked by *ang*.

There is a considerable number of semantic roles which can be expressed on the verb and refer to the *ang*-marked noun phrase. The infix *-um-* marks the item in the *ang*-phrase as an actor (30a). Other roles are expressed as follows: zero suffix for undergoer (30b)), the suffix *-(h)an* for locatives (or datives) (30c), the prefix *ipang-* for instrumentals (30d) and the prefix *i-* for benefactives (30e). The system that establishes the grammatical relation between the predicate and the element in the *ang*-phrase has often been described under the terms of "focus" or "topic", a terminology that is misleading because the function of that system has nothing to do with information structure but rather with reference (cf. Himmelmann 1997: 103 on *ang* as a specific article). For that reason, Schachter (1993) uses the more neutral term of "trigger", which will also be used in this paper. The terms actor-trigger (AT), undergoer trigger (UT), etc. will be used for markers which are related to the actor or the undergoer, respectively.

As Himmelmann (1987) points out, the function of the trigger-morphology runs parallel to a certain type of derivational morphology that determines the orientation of a nominalized action to a participant and its semantic role (cf. the term *Ausrichtung* 'orientation' in Lehmann 1984: 151-152). In the case of the actor trigger, the form is thus oriented to the actor as in agent nouns (*buyer* from *buy; cf. b-um-ili* [buy/AT] in 30a). A similar analysis applies to all the other forms of the trigger. The instrumental trigger, for instance, corresponds to such nominal derivations as *opener* in *tinopener* [the instrument with which a tin is opened]. If this analysis is true, the structure of (30) can be compared to equational constructions cross-linguistically. The first position corresponds to the nominal predicate position that is followed by the *ang*-position (which is not that different from the subject position as suggested by Schachter 1976; cf. Kroeger 1993). The other nominal markers (*ng* and *sa*) are modifiers of the element in the predicate slot. Thus, (30a) can be analysed as 'The man is a buyer of fish at the store', while (30d) corresponds to 'the money is the buying instrument for fish of the man', etc.

As can be seen from examples (31) and (32), the trigger-marked lexemes can occur in both positions of the equational construction, in the predicate position as well as in the *ang*-position. The same also applies to non-trigger-marked lexemes such as *artista* 'actress', *lalaki* 'man' and also *bata* 'child' and *isda* 'fish' in (30):

(31) Tagalog (Schachter and Otanes 1972: 62):

a. *Y-um-aman ang artista.*
get.rich/AT T actress
'The actress got rich.' [The actress is the one who got rich.]

b. *Artista ang y-um-aman.*
actress T get.rich/AT
'The one who got rich is an actress.'

---

[11] In the following example, we find the marker *ng* in a possessive construction:

(i) *lapis ng bata*
pencil POSS child
'the pencil of the child'

(32) Tagalog (Schachter 1985):

a. *Nag-trabaho* *ang* *lalaki*
AT-work T man
'The man worked.' [The man is the one who works.]

b. *Lalaki* *ang* *nag-trabaho*.
man T AT -work
'The one who worked is a man.'

In the following example, we find trigger-marked lexemes in the predicate position and in the *ang*-position:

(33) Tagalog (Himmelmann 2007):
I-u-uwi=nya *ang* *à-alaga-an=nya.*
IT-DUR-return=GEN:3SG T DUR-care.for-LT=GEN:3SG
'He would return the ones he was going to care for.'

Since the slots in an equational construction relate to only one word class, i.e. the noun, the parts-of-speech system of Tagalog can be subsumed under the type of lexically extremely rigid languages, in which there is only one class of content lexemes which can take the nominal predicate position and the *ang*-position of the equational clause. At the same time, Tagalog is also a grammatically rigid language (G_{RIGID}) if one assumes that trigger-marked lexemes are not only clausal in the sentence-initial predicate position but also in the *ang*-position as illustrated in examples such as (31b, *y-um-aman*) and (32b, *nag-trabaho*). In both examples, this assumption is supported if the trigger-marked form is analysed as a predication within a headless-relative clause, i.e., *y-um-aman* 'the one who got rich' in (31b) and *nag-trabaho* 'the one who works' in (31b).

Thus, Tagalog is a type 4 language characterized by the values L_{RIGID} and G_{RIGID}. In spite of this, there are two types of content lexemes in Tagalog from a purely morphological perspective, those that can take trigger morphology and those that cannot. Words like *artista* 'actress' (31), *lalaki* 'man' (32), *bata* 'child' (30) and *isda* 'fish' (30) cannot take trigger morphology, while words like *bili* 'buy' (30), *yaman* 'get rich' (31) and *mag-trabaho* 'work' (32) can. More generally speaking, content lexemes that denote actions can take trigger morphology, while those that denote objects cannot (Himmelmann 2005, 2007). Thus, one can conclude that Tagalog is a language with distinct morpholexical categories but no distinct terminal syntactic categories that are associated with the syntactic slots for nouns and verbs. It shares this property with another type 4 language like Classical Nahuatl (cf. this section) and with Tongan, a type 2 language described by Broschart (1997).

5 Conclusion

The present paper has tried to see what happens to the Typology of Flexibility with its L(exical) and G(rammatical) parameters and their values of FLEXIBLE and RIGID if it is applied to languages that stand for extreme positions within that continuum. A look at the four languages of Late Archaic Chinese, Khmer, Nahuatl and Tagalog revealed the following three areas in which problems may arise:

(i) The G parameter can be problematic in terms of its relevance and of the level of abstraction at which grammatical morphemes are analysed. The data from Late Archaic

Chinese show that there are languages in which the G parameter cannot be addressed and thus turns out to be irrelevant. The majority of Khmer affixes (22 out of 28) have a flexible G parameter from the perspective of their overall functional potential and a rigid G parameter at the level of the individual lexical base with which they are combined.

(ii) Extreme type 4 languages are not necessarily verbs-only languages. Thus, the parts-of-speech hierarchy only operates with omnipredicative languages such as Nahuatl but it fails to account for languages such as Tagalog which consists of nouns that are used in copular constructions.

(iii) Even if a language has only one word class as in the case of extreme type 4 languages, morphology does not necessarily harmonize with these parts-of-speech properties. While Nahuatl has only verbs and Tagalog has only nouns, its morphology clearly distinguishes nouns and verbs and thus does not care about the linking rules that apply from the lexicon to syntax.

To conclude this paper with an outline for further research, I would like to point out that most of the above properties that are critical for the Typology of Flexibility are related to diachronic changes in morphology:

- The irrelevance of the G parameter in Late Archaic Chinese is associated with the loss of morphology at the end of Old Chinese (cf. subsection 2.3).

- In most Philippine languages, there is a clear-cut morphological distinction between finite and non-finite verb forms that enhances the noun/verb distinction in these languages (Ross 2002). Within the family of Philippine languages, the special typological status of Tagalog as a lexically and grammatically rigid language consisting only of nouns can thus be seen as the result of morphological loss. Tagalog has lost the distinction between finite and non-finite verb forms which is crucial for keeping the noun/verb distinction in the L parameter.

- Khmer has an enormous morphological potential for developping a full-fledged morphological system in which affixes get specialised for certain word classes. This potential has not been used due to contact with languages that offered an alternative more syntax-based solution to word formation (cf. section 3).

Observations like these point out that there is a very important diachronic perspective that is often neglected in the discussion of parts of speech. The above three cases show that a lot of the extreme examples presented in this paper may be accounted for in a straighforward way by looking at the morphological changes that produced them.

List of abbreviations

1	First person	N	Noun
2	Second person	NEG	Negation
3	Third person	OBJ	Object
A	Actor	OC	Old Chinese (11^{th}-3^{rd} centuries BC)
AT	Actor Trigger	PDL	Person denoting lexeme
ATTR	Attributive marker	PF	Perfect
BT	Benefactive trigger	PFV	Perfective

DUR	Durative marker	PL	Plural
EQ	Equational marker/copula	POSS	Possession marker
EXCL	Exclamative marker	Q	Question marker
GEN	Genitive	S	Single argument of intransitive verb
IT	Instrumental Trigger	SG	Singular
LNK	Linker	T	Trigger
LOC	Locative particle	U	Undergoer
LT	Locative Trigger	UT	Undergoer trigger
MC	Middle Chinese (1^{st}-10^{th} centuries AD)	V	Verb

References

Bisang, W. (1992). *Das Verb im Chinesischen, Hmong, Vietnamesischen, Thai und Khmer*. Tübingen: Narr.

--- (2001). 'Areality, grammaticalization and language typology. On the explanatory power of functional criteria and the status of Universal Grammar', in W. Bisang (ed.), *Language Typology and Universals*. Berlin: Akademie-Verlag, 175–223.

--- (2008a). 'Precategoriality and syntax-based parts of speech: The case of Late Archaic Chinese', *Studies in Language* 32: 568–589.

--- (2008b). 'Precategoriality and argument structure in Late Archaic Chinese', in J. Leino (ed.), *Constructional reorganization*. Amsterdam – Philadelphia: Benjamins, 55–88.

--- (2008c). 'Underspecification and the noun/verb distinction: Late Archaic Chinese and Khmer', in A. Steube (ed.), *The discourse potential of underspecified structures: Event structures and information structures*. Berlin: Akademie-Verlag, 55–83.

--- (2011). 'Word classes', in: J.-J. Song (ed.), *The Oxford handbook of Language Typology*. Oxford: Oxford University Press, 280–302.

Broschart, J. (1997). 'Why Tongan does it differently: Categorial distinctions in a language without nouns and verbs', *Linguistic Typology* 1: 123–65.

Chen, C. 陳承澤. (1922.) 國文法草創 *Guówénfǎ cǎochuàng* [Beginnings of a Chinese Grammar]. Beijing: Shangwu Yinshuguan.

Clark, E. V. and Clark, H. H. (1979). 'When nouns surface as verbs', *Language* 55: 767–811.

Croft, W. A. (1991). *Syntactic categories and grammatical relations: The cognitive organization of information.* Chicago: Chicago University Press.

--- (2000). 'Parts of speech as typological universals and as language particular categories', in P. M. Vogel and B. Comrie (eds.), *Approaches to the typology of word classes*. Berlin: Mouton de Gruyter, 65–102.

--- (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.

--- (2003). *Typology and universals.* Cambridge: Cambridge University Press. (Second edition).

Enfield, N. J. (2003). *Linguistic epidemiology. Semantics and grammar of language contact in mainland Southeast Asia*. London – New York: Routledge Curzon.

Evans, N. and Osada, T. (2005). 'Mundari: The myth of a language without word classes', *Linguistic Typology* 9: 351–390.

Foley, W. A. and Van Valin, R. D. Jr. (1984). *Functional syntax and universal grammar*. Cambridge: Cambridge University Press.

Gassmann, R. H. (1997). *Grundstrukturen der antikchinesischen Syntax. Eine erklärende Grammatik.* Bern – Frankfurt: Lang.

Gabelentz, G. von der. (1878). 'Beitrag zur Geschichte der chinesischen Grammatiken und zur Lehre von der grammatischen Behandlung der chinesischen Sprache', *Zeitschrift der Deutschen Morgenländischen Gesellschaft (ZDMG)* 32: 601–664.
--- (1881). *Chinesische Grammatik. Mit Ausschluss des niederen Stiles und der heutigen Umgangssprache*. Leipzig: T. O. Weigel.
Graham, A. C. (1957). 'The composition of the Gongsuen Long Tzyy', *Asia Major* 11: 147–183.
Grice, P. H. (1975). 'Logic and conversation', in P. Cole and J. L. Morgan (eds.), *Speech acts*. New York: Academic Press, 41–58.
Haiman, J. (1998). 'Possible origins of infixation in Khmer', *Studies in Language* 22: 597–617.
Henderson, E. (1952). 'The main features of Cambodian pronunciation', *Bulletin of the School of Oriental and African Studies* 14: 149–174.
Hengeveld, K. (1992). *Non-verbal predication: Theory, typology, diachrony*. Berlin: Mouton de Gruyter.
Hengeveld, K., Rijkhoff, J. and Siewierska, A. (2004). 'Parts of speech systems and word order', *Journal of Linguistics* 40: 527–570.
Hengeveld, K. and Rijkhoff, J. 2005. 'Mundari as a flexible language', *Linguistic Typology* 9: 406–431.
Himmelmann, N. P., Jr. (1987). *Morphosyntax und Morphologie — Die Ausrichtungsaffixe im Tagalog.* München: Fink.
--- (1997). *Deiktikon, Artikel, Nominalphrase. Zur Emergenz syntaktischer Strukturen.* Tübingen: Niemeyer.
--- (2005). 'The Austronesian languages of Asia and Madagascar: Typological Characteristics', in K. A. Adelaar and N. P. Himmelmann Jr. (eds.), *The Austronesian Languages of Asia and Madagascar*. London: Routledge, 110–181.
--- (2007). 'Lexical categories and voice in Tagalog', in P. Austin and S. Musgrave (eds.), *Voice and grammatical relations in Austronesian languages*. Stanford: CSLI Publications, 246–293.
Huffman, F. E. (1967). *An outline of Cambodian grammar*. PhD dissertation Cornell University.
--- (1973). 'Thai and Cambodian: A case of syntactic borrowing?', *Journal of the American Oriental Society* 93: 488–509.
Jennner, P. N. and Pou, S. (1980/1981). *A lexicon of Khmer morphology* (Mon-Khmer Studies IX-X). Honolulu: University of Hawaii Press.
Karlgren, B. (1920). 'Le Proto-chinois langue flexionelle', *Journal Asiatique* 15: 205–233.
--- (1957). 'Grammata serica recensa', *Bulletin of the Museum of Far Eastern Antiquities*, 29: 1–332.
Kroeger, P. (1993). *Phrase structure and grammatical relations in Tagalog.* Stanford: California: CSLI Publications, Center for the Study of Language and Information.
Launey, M. (1994). *Une grammaire omniprédicative. Essai sur la morphosyntaxe du Nahuatl classique*. Paris: CNRS Éditions.
Lehmann, C. (1984). *Der Relativsatz*. Tübingen: Gunter Narr.
Levinson, S. C. (2000). *Presumptive meanings. The theory of generalized conversational implicature*. Cambridge, Massachusetts: The MIT Press.
Lewitz, S. (1967). 'La dérivation en cambodgien moderne', *Revue de l'école nationale des langues orientales* 4: 65–84.
--- (1976). 'The infix *-b-* in Khmer', in P. N. Jenner, L. C. Thompson and T. Starosta (eds.), *Austroasiatic Studies II* (Oceanic Linguistics Special Publications No. 13). Honolulu: University of Hawaii Press, 741–760.

Lier, E. van and Rijkhoff, J. (this volume). 'Introduction', in J. Rijkhoff and E. van Lier (eds), *Flexible word classes: a typological study of underspecified parts-of-speech*.
Mei, T.-L. (1989). 'The causative and denominative functions of the *s*-prefix in Old Chinese', *Proceedings of the Second International Conference on Sinology, Academia Sinica, Sections on Linguistics and Paleography*, 33–51.
Mithun, M. (2000). 'Noun and verb in Iroquoian languages: multicategorisation from multiple criteria', in P. M. Vogel and B. Comrie (eds.), *Approaches to the typology of word classes*. Berlin: Mouton de Gruyter, 397–420.
Mosel, U. and Hovdhaugen, E. (1992). *Samoan reference grammar*. Oslo: Scandinavian University Press.
Nacaskul, K. (1978). 'The syllabic and morphological structure of Cambodian words', *Mon-Khmer Studies* 7: 183–200.
Nichols, J., Peterson, D. A. and Barnes, J. 2004. 'Transitivizing and detransitivizing languages'. *Linguistic Typology* 8, 149–211.
Pulleyblank, E. G. (1995). *Outline of Classical Chinese grammar*. Vancouver: University of British Columbia Press.
Ramos, T. V. (1974a). *Tagalog structures*. Honolulu: The University Press of Hawaii.
--- (1974b). *The case system of Tagalog verbs* [Pacific Linguistics, Series B, No. 27]. Canberra: Australian National University.
Rijkhoff, J. (2008). 'On flexible and rigid nouns', *Studies in Language*, 23-3:727–752.
Ross, M. (2002). 'The history and transitivity of western Austronesian voice and voice-marking', in: F. Wouk and M. Ross (eds.), *The history and typology of western Austronesian voice systems*. Canberra: The Australian National University, 17–62.
Sagart, L. 1999. *The roots of Old Chinese*. Amsterdam and Philadelphia: Benjamins.
Sasse, H.-J. (1993a). 'Das Nomen – eine universale Kategorie?', *Sprachtypologie und Universalienforschung* 46: 187–221.
--- (1993b). 'Syntactic categories and subcategories', in J. Jacobs, A. von Stechow, W. Sternefeld and Th. Vennemann (eds.), *Syntax: an international handbook of contemporary research*. Berlin: Mouton de Gruyter, 646–686.
Schachter, P. (1976). 'The subject in Philippine languages: topic, actor, actor-topic, or none of the above?', in C. N. Li (ed.), *Subject and topic*. New York: Academic Press, 493–518.
--- (1985). 'Parts of speech systems', in T. Shopen (ed.), *Language typology and syntactic description, Vol. 1: Clause structure*. Cambridge: Cambridge University Press, 3–61.
--- (1993). 'Tagalog', in J. Jacobs, A. von Stechow, W. Sternefeld and Th. Vennemann (eds.), *Syntax: an international handbook of contemporary research*. Berlin: Mouton de Gruyter. 1418–1430.
Schachter, P. and Otanes, F. T. (1972). *Tagalog reference grammar*. Berkeley, Los Angeles and London: University of California Press.
Schiller, E. (1985). 'Forward into the past, a look at Khmer morphology', in R. Arlene R. K. Zide, D. Magier and E. Schiller (eds.), *Proceedings of the conference on participant roles: South Asia and adjacent areas*. Bloomington: Indiana University Linguistics Club, 82–91.
Van Valin, R. D., Jr. and LaPolla, R. J. (1997). *Syntax. Structure, meaning and function.* Cambridge: Cambridge University Press.
Van Valin, R. D., Jr. (2005). *Exploring the syntax-semantics interface*. Cambridge: Cambridge University Press.
Xu, D. (2006. *Typological change in Chinese syntax*. Oxford: Oxford University Press.
Yang B. 楊伯峻 and Leshi H. 何樂士. 1992. 古漢語語法及其發展 *Gǔ Hànyǔ yǔfǎ jí qí fāzhǎn* [The grammar of Classical Chinese and its development]. Beijing: Yuwen Chubanshe.

Zádrapa, L. 2009. Word-class flexibility in Classical Chinese: Verbal and adverbial uses of nouns. PhD dissertation Univerzita Karlova v Praze/University of Prague.