

# Parts-of-speech systems and word order<sup>1</sup>

KEES HENGEVELD

*University of Amsterdam*

JAN RIJKHOFF

*University of Aarhus*

ANNA SIEWIERSKA

*University of Lancaster*

(Received 5 February 2003; revised 6 January 2004)

This paper argues that the word order possibilities of a language are partly determined by the parts-of-speech system of that language. In languages in which lexical items are specialized for certain functionally defined syntactic slots (e.g. the modifier slot within a noun phrase), the identifiability of these slots is ensured by the nature of the lexical items (e.g. adjectives) themselves. As a result, word order possibilities are relatively unrestricted in these languages. In languages in which lexical items are not specialized for certain syntactic slots, in that these items combine the functions of two or more of the traditional word classes, other strategies have to be invoked to enhance identifiability. In these languages word order constraints are used to make syntactic slots identifiable on the basis of their position within the clause or phrase. Hence the word order possibilities are rather restricted in these languages. Counterexamples to the latter claim all involve cases in which identifiability is ensured by morphological rather than syntactic means. This shows that there is a balanced trade-off between the syntactic, morphological, and lexical structure of a language.

## I. INTRODUCTION

Most typological studies concentrate on syntactic and, to a lesser extent, morphological properties of languages. The structure of the lexicon and its

---

[1] We are grateful to Edith Moravcsik, the editors of *JL* and two anonymous *JL* referees for their comments on an earlier version of this paper.

Abbreviations used in morpheme glosses: 1 = first person, 2 = second person, 3 = third person, ABS = absolutive, ACC = accusative, ANAPH = anaphoric, ART = article, ASP = aspect, AUX = auxiliary, CL = classifier, CO = coordinator, CONN = connector, DAT = dative, DIR = direction, ERG = ergative, FUT = future, GENR = general tense, GER = gerund, INT = interrogative, IT = iterative, LD = locative-directional, LNK = linker, M = masculine, NHUM = non-human, NPAST = non-past, OBJ = object, PAST = past, PERF = perfect, PF = perfective, PL = plural, PM = predicate marker, POSS = possessor, PRES = presentative, PROG = progressive, PRS = present, PUNCT = punctual, RCP = reciprocal, REAL = realis, REL = relativizer, SBJ = subject, SG = singular, SPEC = specific, SS = same subject, STAT = stative, SUB = subordinator, TOP = topic, TR = transitive, VR = verbalizer.

consequences for the language system have received much less attention, a point first made explicit in Lehmann (1990), despite the fact that an increasingly important role is assigned to the lexicon in both formal and functional theories of language.<sup>2</sup> This paper is an attempt to fill part of this gap by investigating the correlations between the parts-of-speech system of a language and some of its morpho-syntactic properties. We thus aim at supplementing the insights provided by syntactic and morphological typology with those provided by lexical typology.<sup>3</sup>

The paper is organized as follows: section 2 presents the language sample on which the research is based. In section 3 we summarize the classification of parts-of-speech systems presented in Hengeveld (1992a) and show how the sample languages fit into this classification. Our definitions of the relevant word order parameters and the classification of the sample languages in terms of these parameters are given in section 4. After these preliminary sections, we are in a position to formulate, in section 5, a number of hypotheses concerning the correlations between parts-of-speech systems and word order properties, which we then systematically test on the languages of the sample. The discussion of our results in section 6 concentrates on the question how our results throw new light on various findings in the earlier typological literature on word order. More specifically, we claim that a number of well-known word order correlations can be restricted to sets of languages sharing the same parts-of-speech system.

## 2. THE SAMPLE

The sample on which the research is based is given in table 1. The languages listed there have been selected in such a way that the sample represents the highest possible degree of genetic, geographic and typological diversity.

In order to meet the genetic criterion, the languages in the sample were selected using the method presented in Rijkhoff et al. (1993). This method aims at creating maximal genetic diversity in the sample and – in this case – has been applied to Ruhlen's (1991)<sup>4</sup> classification of the world's

---

[2] For example, in Chomsky's Minimalist program (Chomsky 1995), which concentrates on invariant ('narrow') syntax, all apparent differences among languages are deemed to be lexical in nature. Thus it would seem that, although the syntactic component as such is trimmed down, more and more information is now delegated to the lexicon.

[3] For an overview of earlier attempts at integration see Plank (1998).

[4] In the second edition of Ruhlen's classification, two major language families – Korean-Japanese-Ainu and Kartvelian – are distinguished that were considered sub-branches of larger phyla (Altaic and Caucasian, respectively) in the first edition of this work. Note that our reliance on Ruhlen's classification does not necessarily commit us to all its details, criticized particularly for postulating very large language families for which the evidence is weak. Since the sampling method used here mitigates the effects on the sample of the postulation of large phyla, this criticism hardly affects the constitution of our sample.

PARTS-OF-SPEECH SYSTEMS AND WORD ORDER

Afro-Asiatic (2)	Chadic (1) Cushitic (1)					Gude Oromo, Boraana Turkish
Altaic (1)						Tuscarora
Amerind (7)	Northern (2)	Almosan-Keresiouan (1) Penutian (1)				Koasati Quechua, Imbabura Guaraní Hixkaryana Pipil Warao
	Andean (1) Equatorial-Tucanoan (1) Ge-Pano-Carib (1) Central Amerind (1) Chibchan-Paezan (1)					Ngalakan Kayardild Nunggubuyu
Australian (3)	Gunwinyuan (1) Pama-Nyungan (1) Nunggubuyu (1)					Nung Tagalog Samoan Paiwan Mundari Miao
Austriac (6)	Austro-Tai (4)	Daic (1) Austronesian (3)	Malayo-Pol. (2)	Western (1) Centr.-E. (1)		Paiwanic (1)
	Austroasiatic (1) Miao-Yao (1)					Miao Basque Burushaski, Hunza Abkhaz Itelmen Tamil West Greenlandic (Etruscan) Nivkh Hurrian Polish Hittite Wambon Alamblak Nasioi Tidore Arapesh, Mountain Georgian Ket Nama Hottentot Japanese (Meroitic) Navaho (Nahali) Babungo Kisi Bambara Krongo Lango Ngiti Berbice Dutch Mandarin Chinese Garó Sumerian Hungarian
Basque (1)						
Burushaski (1)						
Caucasian (1)						
Chukchi-Kamchatkan (1)						
Elamo-Dravidian (1)						
Eskimo-Aleut (1)						
Etruscan (1)						
Nivkh (1)						
Hurrian (1)						
Indo-Hittite (2)	Indo-European (1) Anatolian (1)					
Indo-Pacific (5)	Trans New Guinea (1) Sepik-Ramu (1) East Papuan (1) West Papuan (1) Toricelli (1)					
Kartvelian (1)						
Ket (1)						
Khoisan (1)						
Korean-Japanese-Ainu (1)						
Meroitic (1)						
Na-Dene (1)						
Nahali (1)						
Niger-Kordofanian (4)	Niger-Congo (3)	N.-C. Proper (2)	Central N.-C. (1) West Atlantic (1)			
		Mande (1)				
	Kordofanian (1) East Sudanic (1) Central Sudanic (1)					
Nilo-Saharan (2)						
Pidgins and Creoles (1)						
Sino-Tibetan (2)	Sinitic (1) Tibeto-Karen (1)					
Sumerian (1)						
Uralic-Yukaghir (1)						

Table 1  
The sample

languages. Within the restrictions of the genetic criterion, the sample also represents maximal geographic diversity. Where possible, we have selected languages that are spoken in non-contiguous areas.

The sample represents also maximal typological diversity. Given our specific research question, we made sure that among the languages selected there are representatives of all major parts-of-speech systems as indentified in section 3. In one case the typological criterion forced us to slightly violate the genetic criterion: instead of choosing another language from a Formosan branch of Austriac, we included Tagalog, which now represents the Western branch of the large Malayo-Polynesian family.

	Head	Modifier
Predicate phrase	verb	manner adverb
Referential phrase	noun	adjective

*Figure 1*  
Lexemes and syntactic slots

The applicability of the criteria mentioned above is dependent upon the availability of adequate language descriptions. For three out of the 53 languages that were selected according to the genetic criterion, data are insufficient or lacking. The languages concerned are the three isolates, Etruscan, Meroitic and Nahali. Thus the actual sample contains 50 languages.

### 3. PARTS OF SPEECH

#### 3.1 *The syntactic approach*

Hengeveld (1992a, b) presents a classification of parts-of-speech systems and demonstrates the predictive value of this classification for the typology of systems of non-verbal predication. The classification is syntactically-based, that is, lexeme classes are defined in terms of the syntactic slots they may occupy.

The four syntactic slots investigated are head and modifier of a referential phrase (i.e. noun phrase), and head and modifier of a predicate phrase. Heads are obligatory slots and modifiers are optional slots within phrases. The distinguishing uses of classes of lexemes may be represented as in figure 1.

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 we restrict ourselves to manner adverbs. We exclude other classes of adverbs, such as temporal and spatial ones, which do not modify the head of the predicate phrase, but rather modify the sentence as a whole.

In contrast with many earlier approaches, Bhat (1994) being a clear exception, lexemes are thus defined primarily in terms of their non-predicative uses. Verbs are characterized by the fact that they have no non-predicative uses, i.e. they can be used predicatively only. Non-verbal lexemes, on the other hand, may have additional, predicative uses, but their defining use is a non-predicative one.

3.2 *Differentiated, flexible and rigid systems*

Some languages have separate lexeme classes, i.e. noun, adjective, verb and manner adverb, to fill each of the syntactic slots under investigation. English<sup>5</sup> is a case in point:

- (1) The little<sub>A</sub> girl<sub>N</sub> danced<sub>V</sub> beautifully<sub>MAdv</sub>.

A language such as English can be said to have a DIFFERENTIATED PARTS-OF-SPEECH SYSTEM.

Other languages do not have separate lexical classes for each of the syntactic slots. These languages are of two types, as illustrated in (2)–(4).

*Warao* (Romero-Figeroa 1997: 49f.)

- (2) (a) **yakera**  
beauty  
'beauty'  
(b) Hiaka **yakera** auka saba tai nisa-n-a-e.  
garment beauty daughter for she buy-SG-PUNCT-PAST  
'She bought a beautiful dress for her daughter.'

*Garo* (Burling 1961: 27, 33)

- (3) (a) **Da'r**-aŋ-gen.  
big-IT-FUT  
'It will get big.'  
(b) **da'r**-gipa mande  
big-REL man  
'the big man'  
(4) (a) **Ca'**-gen-ma.  
eat-FUT-INT  
'Will you eat?'  
(b) **ca'**-gipa mande  
eat-REL man  
'the man who eats'

In Warao the translational equivalents of English adjectives and nouns can be used both as heads of referential phrases and as modifiers within referential phrases. An example is the lexeme *yakera* 'beauty' in (2), which is used as the head of a referential phrase in (2a) and as a modifier within a referential phrase in (2b). In Garo, on the other hand, the translational equivalents of English adjectives and verbs, e.g. *da'r* 'big' in (3) and *ca'* 'eat' in (4), seem at first sight to be used both as heads of predicate phrases,

[5] In English there are of course many cases of zero conversion between the classes of noun and verb. These cases will be discussed in 3.4.

as in (3a) and (4a), and as modifiers within referential phrases, as in (3b) and (4b).

Facts like these have led many authors (Schachter 1985, Wetzer 1996, Stassen 1997) to conclude that there are ‘adjectival-noun’ languages, such as Warao, and ‘adjectival-verb’ languages, such as Garo, as if these were two equivalent options. This approach overlooks an important difference between the two types of language. In Warao there is indeed one class of lexemes which may occur as head and as modifier within a referential phrase, i.e. members of this class may occupy two different syntactic slots. But in Garo this is not the case. In order to use the lexemes involved as modifiers within a referential phrase, they have to be relativized, which shows that the modifier slot is not occupied by a lexical unit but by a syntactic unit, in this case a relative clause. Within this clause the lexeme functions as the head of a predicate phrase, i.e. as a verb, just as it does in main clauses.

Thus, Warao has one class of lexemes which may occupy the syntactic slots of head and modifier within a referential phrase, whereas Garo lacks a class of lexical elements which may occupy the modifier slot in a referential phrase, and therefore this language has to resort to alternative, non-lexical strategies. A language such as Warao can be said to have a FLEXIBLE PARTS-OF-SPEECH SYSTEM, whereas a language such as Garo can be said to have a RIGID PARTS-OF-SPEECH SYSTEM.

This difference extends to the class of manner adverbs as well. Consider the following examples:

*Warao* (Romero-Figeroa 1997: 119)

- (5) *Oko kuana yaota-te arone yakera nahoro-te ...*  
 we hardness work-NPAST although beauty eat-NPAST  
 ‘Although we work hard and eat well, ... .’

*Garo* (Burling 1961: 29)

- (6) (a) *Rak-e dok-aha.*  
 strong-SUB hit-PAST  
 ‘He hit hard.’  
 (b) *Bia gar-e kat-an-aha.*  
 3.SG throw-SUB run-IT-PAST  
 ‘Throwing, he ran away.’

In Warao the modifier slot in a predicate phrase may be occupied by lexical elements which also occur as heads and modifiers within referential phrases, as is evident from a comparison of (5) with (2). In Garo the manner expression is a verb with the suffix *-e*, which is a subordinating suffix with a manner meaning that can be added to any kind of verb, as (6a), which contains the verb *rak-* ‘(be) strong’, shows. So again Warao shows flexibility in its parts-of-speech system, whereas Garo shows rigidity.

Language	Head of predicate phrase	Head of referential phrase	Modifier of referential phrase	Modifier of predicate phrase
Warao	verb	non-verb		
English	verb	noun	adjective	manner adverb
Garo	verb	noun	–	–

Figure 2  
Flexible, differentiated and rigid languages 1

The differences among Warao, English and Garo are represented in figure 2. This figure shows that Warao has a class of lexical elements, NON-VERBS, which combine the functions that nouns, adjectives and manner adverbs have in languages with a differentiated parts-of-speech system, whereas in Garo lexical elements which fulfil the functions of adjectives and manner adverbs are simply lacking.

### 3.3 *The parts-of-speech hierarchy*

The degree of flexibility/rigidity displayed within a parts-of-speech system differs from language to language. Interestingly, the combinations of syntactic possibilities for a single lexical class in flexible languages and the lack of lexical classes for certain syntactic slots in rigid languages are not random. The variation can be described in terms of the hierarchy in (7).

(7) *Parts-of-speech hierarchy*

Head of predicate phrase	>	Head of referential phrase	>	Modifier of referential phrase	>	Modifier of predicate phrase
--------------------------------	---	----------------------------------	---	--------------------------------------	---	------------------------------------

The parts-of-speech hierarchy should be interpreted in the following way: the more to the left a certain syntactic slot is positioned in the hierarchy, the more likely it is for a language to have a separate class of lexemes for that syntactic slot.<sup>6</sup> This means that if a rigid language lacks a class of lexemes that may be used as modifiers within a referential phrase, i.e. adjectives, it will also lack a class of lexemes that may be used as modifiers within predicate phrases, i.e. manner adverbs. And if a flexible language has a class of lexemes that may be used as heads and modifiers in referential phrases, these lexemes will also be used as modifiers within predicate phrases, i.e. this language has a class of non-verbs. A number of parts-of-speech systems are predicted by this hierarchy, the differences between them depending on the degree of flexibility/rigidity displayed.

[6] Each of the steps in the hierarchy correlates with a number of other features of a language. See, for example, Rijkhoff (2002: 141–145; 2003), who proposes necessary conditions for the occurrence of distinct classes of verbs, nouns and adjectives in a language.

A lower degree of flexibility than that observed in Warao can be found in Ngiti, and a lower degree of rigidity than that observed in Garo can be found in Wambon:

*Ngiti* (Kutsch Lojenga 1993: 338)

- (8) (a) Ngbángba nítdù                    **ísó**    ànò.  
 child                    3.carry.PF.PRS light load  
 ‘The child carried a light load.’  
 (b) **Isó** ngbángba nítdù                    ànò.  
 light child                    3.carry.PF.PRS load  
 ‘The child carried the load easily.’

*Wambon* (de Vries & de Vries-Wiersma 1992: 19, 90)

- (9) (a) Jakhov-e **matet-mo** ka-lembo?  
 they-CONN good-VR.SS go-3.PL.PAST  
 ‘Did they travel well?’  
 (b) Nggerkaji-ke-levambo-n-o ko-khe-n-o  
 saw-VR-I.PL.PAST-TR-CO go-3.SG.PRS-TR-CO  
**kutip-ke-lo nda-tulo** la-levambo.  
 night-VR-SS come-ascend.SS sleep-I.PL.PAST  
 ‘We continued sawing until the night fell and we went up and slept’.

In Ngiti the word *ísó* ‘light, easily’ can be used as a modifier within a referential phrase, as in (8a), and as a modifier within a predicate phrase, as in (8b). Wambon, on the other hand, lacks a class of manner adverbs. It uses medial verb constructions to create manner expressions. This can be seen in (9a), where the verb form *matetmo*, a verbalized form of the adjective *matet* ‘good’, is the head of the predicate phrase of the medial verb construction, which itself modifies the main verb. The medial verb construction is a cosubordinating strategy widely used in the language, as illustrated in (9b). Thus, Ngiti combines the functions of adjectives and manner adverbs in a single class of lexical elements, MODIFIERS, whereas Wambon lacks a class of manner adverbs and compensates for this absence by means of a syntactic solution. The differences between Ngiti and Wambon are shown in figure 3.

A higher degree of flexibility than that observed in Warao and a higher degree of rigidity than that observed in Garo can also be found. The flexible extreme defined by the parts-of-speech hierarchy is a language with no lexical specialization at all. An example in our sample is Samoan. In this language any lexical item can basically be used in any syntactic slot, the only restriction being semantic compatibility. The following examples illustrate the flexibility of lexical items in Samoan in three different ways. First, a particularly striking property of Samoan is that the translational equivalents of an English noun can be used not only as the head of a



Language	Head of predicate phrase	Head of referential phrase	Modifier of referential phrase	Modifier of predicate phrase
Ngiti	verb	noun	modifier	
English	verb	noun	adjective	manner adverb
Wambon	verb	noun	adjective	–

Figure 3

Flexible, differentiated and rigid languages 2

referential phrase but also as the head of a predicate phrase, as in (10), whereas the translational equivalent of an English verb can not only be used as the head of a predicate phrase but also as the head of a referential phrase, as in (11).

*Samoan* (Mosel & Hovdhaugen 1992: 80, 73)

(10) (a) 'Ua mālosi le **lā**.

PERF strong ART sun

'The sun is strong.' ('The sun strongs.')

(b) 'Ua **lā** le aso.

PERF sun ART day

'The day is sunny.' ('The day suns.')

(11) (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'

Secondly, the translational equivalent of an English noun can be used as a modifier within a referential phrase, as in (12), or as a modifier within a predicate phrase, as in (13). Note that the absence of an article with the modifying lexeme shows that the lexeme is used in a non-referential slot.

*Samoan* (Mosel & Hovdhaugen 1992: 305, 394)

(12) le fale **ta'avale**

ART house car

'the garage' ('the car house')

(13) 'Ua ma'i **misela** le tama.

PERF sick measles ART boy

'The boy has got the measles.' ('The boy sick measleswise.')

Thirdly, the translational equivalent of an English verb can be used as a modifier within a referential phrase, as in (14), and as a modifier within a predicate phrase, as in (15). Here the absence of a tense-aspect particle with the modifying lexeme shows that the lexeme is used non-predicatively.

*Samoan* (Mosel & Hovdhaugen 1992: 294, 397)

- (14) le mea **ta'alo**  
 ART thing play  
 'the toy' ('the play thing')
- (15) ... 'olo'o moe **taagulu** ai Vesi  
 ... PROGR sleep snore ANAPH Vesi  
 '... where Vesi was sleeping snoring.'

The rigid extreme defined by the hierarchy in (7) would be a language that has verbs only. Such a language is not attested in our sample, but a sample language that helps to understand what an extremely rigid language might look like is Tuscarora.<sup>7</sup> This language does have a class of morphological nouns (see Mithun 2000). In many cases, however, in this language one has to use a full predication in order to render the meaning of an English noun. Examples (16) and (17) illustrate this phenomenon.

*Tuscarora* (Mithun 1976: 35, 30)

- (16) ra-kwatihs  
 M-young  
 'boy' ('He is young.')
- (17) ka-teskr-ahs  
 NHUM-stink-ASP  
 'goat' ('It stinks.')

Thus, many lexemes in Tuscarora have a predicative use only and should therefore be classified as verbs. As a consequence, what in many other languages would be a single predication may in Tuscarora show up as a set of appositional predications, as shown in (18):

*Tuscarora* (Mithun 1976: 32)

- (18) ra-kwatihs wa-hr-atkahto-ʔ ka-teskr-ahs  
 M-young PAST-M-look.at-ASP NHUM-stink-ASP  
 'The boy looked at the goat.' ('He is young, he looked at it, it stinks.')

In a similar way appositional predications are used instead of adjectival and adverbial modifiers, as shown in (19) and (20).

*Tuscarora* (Mithun 1976: 234, 256)

- (19) tá:ko:θ yaw-vhey-vʔ  
 cat NHUM.OBJ-die-PF  
 'the dead cat' ('the cat, it has died/is dead')  
 or 'The cat has died/is dead.'

[7] An extensive and insightful discussion of the verbal orientation of Iroquoian languages may be found in Sasse (1988).

Language	Head of predicate phrase	Head of referential phrase	Modifier of referential phrase	Modifier of predicate phrase
Samoan	contentive			
English	verb	noun	adjective	manner adverb
(unattested)	verb	–	–	–

Figure 4  
Flexible, differentiated and rigid languages 3

Parts-of-speech system (PoS)	Head of predicate phrase	Head of referential phrase	Modifier of referential phrase	Modifier of predicate 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 5  
Parts-of-speech systems

- (20) yo-hstoreʔ      wa-hr-o-horvh-ʔ  
 NHUM.OBJ-fast.PF    PAST-M-OBJ-grow-ASP  
 ‘He grew fast.’ (‘It is fast, he grew.’)

Thus, Tuscarora lacks adjectives and manner adverbs, and often uses verbs rather than nouns, whereas Samoan combines the functions of verbs, nouns, adjectives and manner adverbs in a single class of lexemes. Figure 4 shows the differences between a purely flexible language such as Samoan and an unattested purely rigid language.

### 3.4 *Parts-of-speech systems*

The combination of the threefold distinction between flexible, differentiated and rigid languages with the parts-of-speech hierarchy results in the classification of parts-of-speech systems presented in figure 5.

In interpreting this classification the following points have to be taken into account:

- (i) In classifying languages in terms of the parts-of-speech hierarchy in (7), we have taken both basic and derived lexemes into consideration. Thus, English is classified as a language with a class of manner adverbs, despite the fact that virtually all these adverbs are derived.

- (ii) In order for a language to qualify as flexible, it has to show SYSTEMATIC flexibility, both semantically and syntactically, for an entire class of lexemes. Thus, English is not classified as a flexible language, despite the fact that in this language there are many cases of zero conversion between the classes of noun and verb.<sup>8</sup>
- (iii) The use of lexemes in syntactic slots can only be evaluated to the extent that languages actually have these slots available. For instance, Himmelmann (forthcoming a) shows that Tagalog does not have a slot for manner modifiers. Instead, it uses a complement-taking predicate describing the manner in which the event described in the subordinate clause takes place. In cases like these, the parts-of-speech system of a language has to be determined on the basis of the behaviour of lexeme classes with respect to their distribution across the remaining slots.

### 3.5 *Intermediate systems*

Some languages do not fit into the basic classification of parts-of-speech systems given in figure 5. All these languages occupy intermediate positions between contiguous types in figure 5. They can thus be said to have INTERMEDIATE PARTS-OF-SPEECH SYSTEMS. The criteria for classifying a language as having an intermediate parts-of-speech system are different for flexible and rigid languages.

In order for a flexible language to qualify as having an intermediate parts-of-speech system, its lexeme classes should be compatible with two contiguous systems within the hierarchy at the same time. This situation may obtain, for instance, when derived lexemes have fewer functional possibilities than basic lexemes within a language. Thus, Mundari is a fully flexible type 1 language if its basic lexemes are taken into consideration, but it also has a derivational process that produces lexemes that can be used in all slots apart from the predicate slot, a type 2 feature. In order to account for these facts, Mundari is classified as a type 1/2 language. Another example of a language with an intermediate system is Lango, which displays a lower degree of flexibility than Mundari. In this language, there is a large open class of lexemes that may be used as modifiers within referential phrases and as modifiers within predicate phrases. On the basis of this fact, it should be classified as a type 3 language. In addition, however, it has a large open class of manner adverbs, a feature of a type 4 language. Therefore, Lango is classified as a type 3/4 language.

A rigid language is classified as having an intermediate parts-of-speech system when the last class of lexemes on the hierarchy that is relevant for that language is a small closed class of items. Thus, Pipil is a language that

---

[8] See Vogel (2000) on the classification of English in terms of the typology of parts-of-speech systems used in this article.

has an open class of adjectives but a restricted set of manner adverbs borrowed from Spanish, and is therefore classified as a type 4/5 language. Tamil, a language with a higher degree of rigidity, has no manner adverbs. It does have adjectives, but only a very limited number of them in a small closed class. For this reason, it is classified as a type 5/6 language.

### 3.6 *Classification of sample languages*

Including the intermediate types discussed in the previous paragraph, languages may be assigned to one of 13 types. Our sample languages fit into these types in the way indicated in table 2. In this table the languages are ordered from extremely flexible (type 1) to extremely rigid (type 7).

PoS	Languages
1	Samoan, Tagalog
1/2	Mundari
2	Hurian, Quechua, Warao
2/3	Turkish
3	Ket, Miao, Ngiti, Tidore
3/4	Lango
4	Abkhaz, Arapesh, Babungo, Bambara, Basque, Burushaski, Georgian, Hittite, Hungarian, Itelmen, Japanese, Nama, Ngalakan, Polish
4/5	Koasati, Nasioi, Paiwan, Pipil, Sumerian
5	Alamblak, Berbice Dutch, Guaraní, Kayardild, Kisi, Oromo, Wambon
5/6	Garo, Gude, Mandarin Chinese, Nung, Tamil, West Greenlandic
6	Hixkaryana, Krongo, Navaho, Nivkh, Nunggubuyu
6/7	Tuscarora
7	–

*Table 2*  
Parts-of-speech systems of sample languages

### 3.7 *The semantics of flexible lexemes*

At this point it might be appropriate to briefly discuss the semantics of flexible lexemes (contentives, non-verbs, modifiers), since this appears to have led to confusion among some of the reviewers of our typology of parts-of-speech systems. For example, both Evans (2000: 729) and Croft

(2001: 70–75) use terms such as ‘polysemy’ and ‘semantic shift’ when they discuss the semantics of flexible lexemes, which they interpret as having distinct but related senses. In our analysis, however, flexible lexemes are SEMANTICALLY VAGUE, i.e. they are monosemous.

A well-known example of semantic vagueness is English ‘cousin’, which is vague with respect to the distinction ‘male cousin’ and ‘female cousin’, or, as Cruse (1986: 51) puts it: ‘the two meanings “male cousin” and “female cousin” are both associated with the same lexical unit cousin, whose meaning is more general than either; they therefore do not represent distinct senses of cousin’. This can be demonstrated with a test which involves anaphoric reference with *so* (Saeed 2003: 61):

(21) Sarah is my cousin, and so is Peter.

If a noun has a vague meaning, as in the case of *cousin* in (21), its sense allows for different specifications in the case of anaphoric reference with *so*. In this article we are concerned with a more fundamental and abstract kind of vagueness, categorial vagueness, which holds across distinct lexical categories (verbs, nouns, etc.) and is therefore more difficult to identify.

The idea that flexible lexemes are semantically vague rather than polysemous finds initial support in what has been written about the semantic nature of flexible lexemes in the grammars of the languages in question. Thus, Mosel & Hovdhaugen (1992: 73) write about Samoan that ‘categorization of full words is not given a priori in the lexicon. It is only their actual occurrence in a particular environment which gives them the status of a verb or a noun’. In a similar vein, Hoffmann (1903: xxxii) notes the ‘functional elasticity’ of Mundari lexemes.

From a typological perspective, semantic vagueness (also known as ‘underspecification’, ‘non-determinedness’, ‘indeterminacy’, ‘generality’) is not an uncommon phenomenon. For example, nouns in many (perhaps even most) languages across the globe are transnumeral in that the same (unmarked) noun may be used to refer to a single individual or to a plurality of entities (Rijkhoff 2002: chapters 2 and 5). Vagueness is also attested in members of other word classes, such as verbs. For example, McGregor (2002: 54–87) shows that in Northwest Australian languages many uninflecting verbs (‘preverbs’) are essentially vague with respect to certain Aktionsart distinctions, valency, and reflexivity; Bisang (1996: 520) mentions the high degree of indeterminateness of nouns and verbs in the languages of East and Mainland South East Asia.

Wilkins’ (2000) account of noun semantics in Arrernte is particularly instructive in the present context. In Arrernte and other Australian languages NPs consist of a specific noun preceded by a generic noun.<sup>9</sup> Together these

---

[9] On this construction see also Dixon (1980: 102f.).

nouns constitute what Wilkins calls ‘classifier constructions’ and he argues that these constructions serve to determine which properties of an entity are relevant from the perspective of the current discourse context (Wilkins 2000: 148). Consider the following examples:

- (22) *Arrernte* (Wilkins 2000: 179–200)
- (a) *yerre arlkerrke*  
ant meat-ant
  - (b) *awelye arlkerrke*  
traditional medicine meat-ant
  - (c) *apmere arlkerrke*  
socially.relevant.place meat-ant

Wilkins (2000: 179–200) provides detailed, monosemous definitions of both the specific noun *arkerrke* ‘meat-ant’ and the three generic nouns *yerre* ‘ant’, *awelye* ‘traditional medicine’ and *apmere* ‘socially relevant place’, and goes on to argue that each combination of a generic noun (*yerre*, *awelye*, *apmere*) and the specific noun *arkerrke* highlights a different set of culture-specific knowledge structures associated with *arkerrke* and that as a consequence the other knowledge structures associated with *arkerrke* are downplayed and backgrounded. For example, according to Wilkins (2000: 192), the semantic effect of the combination *awelye arlkerrke* in (22b) could be paraphrased as follows: ‘In using the classifier construction *awelye arlkerrke* “traditional medicine meat-ant” in a noun phrase, the speaker wants the addressee to think about the referent of the noun phrase from the point of view of its having the properties of an *arkerrke* “meat-ant” but whose discourse properties at the current point in the discourse are its properties as an *awelye* “traditional medicine”’.

We propose a similar approach to the meaning and the function of flexible lexemes: each flexible lexeme has a single (vague) sense. By placing the flexible lexeme in a particular syntactic slot or by providing it with certain morphological markers, the speaker highlights those meaning components of the flexible lexeme that are relevant for a certain lexical (verbal, nominal, etc.) function. Thus we contend that the meaning of a flexible lexeme always remains the same, and that morpho-syntactic and other contextual clues signal to the addressee how to interpret this lexeme in an actual utterance. In other words, it is the use of a vague lexeme in a certain context (an actual linguistic expression) that brings out certain parts of its meaning, giving the category-neutral lexeme a particular categorial (verbal, nominal, etc.) flavour.

Note, finally, that even though a flexible lexeme is not semantically ambiguous, the employment of a flexible lexeme may lead to functional ambiguity, as when the context does not provide sufficient clues regarding the way it is used in the actual linguistic expression. We will return to the issue of functional ambiguity in section 5.1 below.

#### 4. WORD ORDER AND MORPHOLOGICAL MARKERS

##### 4.1 *Introduction*

Ever since Greenberg (1966) the domain of word order has constituted one of the major testing grounds for the predictive power of potential typological parameters. We too will document the merits of our classification of parts-of-speech systems by showing how it interacts with word order phenomena and how it helps to refine some of the word order correlations established by Greenberg and others. In section 5, several word order predictions stemming from our classification of parts-of-speech systems will be presented. These predictions will involve the basic and variable orders of the predicate and its arguments, of the head and modifier in the referential phrase, and of the head and modifier in the predicate phrase. But first some comments on our classification of these linearization patterns and their distribution among the languages in the sample are in order.

##### 4.2 *Clausal order and morphological marking*

Given the existence of languages without a true class of verbs, we classify clausal word order in terms of the location of predicates, rather than of verbs, relative to their arguments. Though main, positive, declarative clauses with two overt referential phrases are not necessarily highly frequent in languages, our classification is based on the order obtaining in such clauses, as in most typological classifications of clausal word order. However, rather than using the traditional Greenbergian six-way typology of SVO, SOV, VSO, VOS, OVS and OSV, we use a three-way typology of predicate-initial, predicate-medial and predicate-final clausal orders. This three-way typology based on the position of the predicate has two major advantages over the six-way typology. First of all, since languages exhibit more variation in the location of the transitive arguments relative to each other than in the location of the predicate phrase relative to both of the arguments (see e.g. Steele 1978, Siewierska 1998), it is easier to assign a basic order to languages in terms of the three-way typology than in terms of the six-way typology. Secondly, of the three logically possible locations of the predicate phrase in transitive clauses, the predicate-initial and predicate-final positions are directly applicable to the classification of clause ordering in intransitive clauses. Consequently, the predicate-based typology allows one to capture any consistencies in the ordering of transitive and intransitive predicates that may obtain in a transparent manner.

Our major criterion for assigning a basic order is statistical frequency. In languages exhibiting considerable word order variation, we assign a unique basic order only if one of the word order patterns is at least twice as common as any other order, following Dryer (1997). If there is no such distribution, no basic order is assigned.



## PARTS-OF-SPEECH SYSTEMS AND WORD ORDER

Language	Clause			Referential phrase			Predicate phrase		
	1 BWO	2 FixSP	3 Morph	4 BWO	5 FixSP	6 Morph	7 BWO	8 FixSP	9 Morph
Abkhaz	3	-	-	HM	-	-	MH	+	-
Alamblak	3	-	-	MHM	-	-	irr	irr	irr
Arapesh	2	+	irr	MHM	-	-	MHM	-	-
Babungo	2	+	irr	HM	+	-	HM	+	-
Bambara	3	+	irr	HM	+	+	HM	+	-
Basque	3	-	-	HM	-	-	MH	+	-
Berberice Dutch	2	+	irr	MH	+	-	irr	irr	irr
Burushaski	3	-	-	MH	-	-	MH	-	-
Garó	3	+	irr	HM	+	-	irr	irr	irr
Georgian	2/3	-	-	MH	+	-	MH	+	-
Guaraní	2	-	-	HM	+	-	irr	irr	irr
Gude	1/2	-	-	MH	+	-	irr	irr	irr
Hittite	3	+	irr	MHM	-	-	MH	+	-
Hixkaryana	2	-	-	irr	irr	irr	irr	irr	irr
Hungarian	2/3	-	-	MH	+	-	MH	+	-
Hurrian	3	+	irr	MH	+	-	MH	+	-
Itelmen	2/3	+	irr	MHM	-	-	MH	+	-
Japanese	3	+	irr	MH	+	-	MH	+	-
Kayardild	1/2/3	-	-	MH	-	-	irr	irr	irr
Ket	3	+	irr	MH	+	-	MH	+	-
Kisi	2/3	+	irr	HM	+	+	irr	irr	irr
Koasati	3	-	-	HM	+	-	MHM	-	-
Krongo	1/2	-	-	irr	irr	irr	irr	irr	irr
Lango	2	+	irr	HM	+	+	HM	+	+
Mandarin Chinese	2	+	irr	MH	+	-	irr	irr	irr
Miao	2	+	irr	HM	+	-	HM	+	-
Mundari	3	+	irr	MH	+	-	MH	+	-
Nama	3	-	-	MH	+	-	MH	+	-
Nasioi	3	-	-	HM	-	-	MH	+	-
Navaho	3	+	irr	irr	irr	irr	irr	irr	irr
Ngalakan	1/2	-	-	HM	-	-	MH	+	-
Ngiti	2/3	+	irr	MH	+	-	MHM	-	-
Nivkh	3	+	irr	irr	irr	irr	irr	irr	irr
Nung	2	-	-	HM	+	-	irr	irr	irr
Nunggubuyu	1/2/3	-	-	irr	irr	irr	irr	irr	irr
Oromo	3	+	irr	HM	+	-	irr	irr	irr
Paiwan	1	-	+	HM	-	+	HM	+	+
Pipil	1	-	-	MH	-	-	MH	-	-
Polish	2	-	-	MHM	-	-	MHM	-	-
Quechua	3	-	+	MH	+	-	MH	+	-
Samoan	1	-	+	HM	+	-	HM	+	-
Sumerian	3	+	irr	HM	-	-	MH	+	-
Tagalog	1	-	+	MHM	-	+	irr	irr	irr
Tamil	3	-	-	MH	+	-	irr	irr	irr
Tidore	2	+	irr	HM	+	-	HM	+	-
Turkish	3	+	irr	MH	+	-	MH	+	-
Tuscarora	2	-	-	irr	irr	irr	irr	irr	irr
Wambon	3	+	irr	HM	+	-	irr	irr	irr
Warao	3	+	irr	HM	+	-	irr	irr	irr
West Greenlandic	3	-	-	HM	+	-	MH	+	+

Table 3  
Word order data

A complete overview of our word order data is given in table 3. In this table 'irr' indicates that the parameter under investigation is irrelevant to the language concerned, in view of the absence of one or more of the parts of speech under investigation. The distribution of basic (BWO) clausal orders among the languages in our sample is shown in column 1, in which 1 = predicate-initial order, 2 = predicate-medial order, and 3 = predicate-final order. The languages with double classifications, either 1/2 or 2/3, are languages in which the order is dependent on tense/aspect (e.g. Gude, Kisi,

Krongo, Ngiti) or languages which display two patterns with a comparable degree of frequency due to pragmatic factors (e.g. Hungarian,<sup>10</sup> Georgian<sup>11</sup>). Those with a triple classification are the languages which lack a basic order.

The percentage of predicate-initial languages (13.4%)<sup>12</sup> in our sample is lower, and that of predicate-final languages (51.3%) higher, than in other samples of comparable size, such as those of Steele (1978) or of Gilligan and of Perkins discussed in Hawkins & Gilligan (1988), while the percentage of predicate-medial languages (35.3%) is roughly the same.<sup>13</sup> The differences are partly due to our assignment of double orders for languages in which order is dependent on tense/aspect or on pragmatic factors. Otherwise, the distribution of clausal orders in our sample is in line with what previous studies would lead one to expect: predicate-final languages clearly outnumber predicate-medial ones, and these in turn outnumber predicate-initial languages.

In addition to basic clausal order, table 3 also lists variation in the ordering of the referential phrase in subject function and the predicate phrase (cf. section 3.1 above). For the purposes of this article the subject of an intransitive sentence is the single argument occurring within that sentence; and the subject of a transitive sentence is the constituent that shows the same syntactic behaviour as that of the single argument of an intransitive sentence. In considering whether a language exhibits any variation in the ordering of subject and predicate phrase, again only main positive declarative clauses are taken into account. Only variations in the ordering of subject and predicate phrase independent of the complexity of the subject and independent of the necessary presence of constituents other than the predicate and its arguments are listed. The data on the variation in the ordering of subject (S) and

[10] See Behrens (1982: 161).

[11] See Vogt (1971: 220–224).

[12] Languages exhibiting more than one basic order are assigned proportionally to the relevant types, e.g. a language having two basic word orders counts as 0.5 for each of the types that it exhibits.

[13] The figures are as follows:

WORD ORDER TYPE:	3	2	1	ALL
<i>This article</i>				
Number of languages	27.17	16.66	6.17	50
%	54.33%	33.33%	12.33%	100%
<i>Hawkins &amp; Gilligan (1988)</i>				
Number of languages	19	20	11	50
%	38%	40%	22%	100%
<i>Steele (1978)</i>				
Number of languages	30	20	13	63
%	47%	32%	21%	100%

predicate (P) are given in table 3 in column 2. In this column, a + indicates that there is a fixed order of S and P with respect to each other, i.e. the subject either always precedes or always follows the predicate, and a – indicates that there is variation in the ordering of S and P.

Since we are interested in the interaction between syntax and morphology as two potentially competing disambiguating strategies in flexible languages, we indicate in column 3 whether there is a special morphological procedure which identifies a change in the basic order of subject and predicate. In this column, a + indicates that there is such a morphological procedure, and a – that there isn't.

#### 4.3 *Order and morphological markers in the referential phrase*

At the level of the referential phrase we consider the order of heads (Hs) and lexical modifiers (Ms) only. Note that we deliberately do not use the terms 'noun' and 'adjective' here, since these were shown in section 3 to be non-universal categories. Depending on the nature of the parts-of-speech system, the head slot of the referential phrase may be filled with CONTENTIVES (in types 1–1/2),<sup>14</sup> NON-VERBS (in types 1/2–2/3), or NOUNS (in types 2/3–6/7). The lexical modifier slot of the referential phrase may be filled with CONTENTIVES (in types 1–1/2), NON-VERBS (in types 1/2–2/3), MODIFIERS (in types 2/3–3/4) or ADJECTIVES (in types 3/4–5/6).

As in the case of clausal order, we assign a basic order for heads and modifiers in referential phrases on the basis of frequency. Both basic orders are assigned to languages in which neither of the orders can be said to be more frequent than the other. Languages in which both orders are possible but for which the information available does not allow us to make a decision on frequency are included in the latter category. The data are listed in column 4 in table 3, where HM = head-modifier basic order, MH = modifier-head basic order, and MHM = both basic orders.

We also list possible variation in the order of head and modifier with respect to each other. In column 5 in table 3, a + indicates that there is a fixed order of H and M with respect to each other, i.e. the modifier either always precedes or always follows the head, and a – indicates that there is variation in the ordering of head and modifier. Of course all languages which have been assigned two basic orders in column 4 are identified as languages with variable order in column 5.

Column 6 indicates whether the language in question uses morphological markers that identify a modifier within a term phrase, such as an attributive particle, or an agreement marker that is exclusively used with attributive

[14] We use notations like '1–2/3' to refer to the segment of the hierarchy from type 1 to type 2/3, and thus including types 1/2 and 2.

modifiers.<sup>15</sup> In this column a + indicates that there is such a morphological marker, a – that there isn't.

#### 4.4 *Order and morphological markers in the predicate phrase*

The data on order in the predicate phrase are comparable to those on order in the referential phrase. Column 7 in table 3 lists the basic order of heads (Hs) and lexical modifiers (Ms) within predicate phrases. Here we avoid the terms 'verb' and 'adverb', since, depending on the parts-of-speech system, the head slot of the predicate phrase may be filled with CONTENTIVES (in types 1–1/2) or VERBS (in types 1/2–7), and the lexical modifier slot of the predicate phrase may be filled with CONTENTIVES (in types 1–1/2), NON-VERBS (in types 1/2–2/3), MODIFIERS (in types 2/3–3/4) or MANNER ADVERBS (in types 3/4–4/5). The basic orders listed are again based on relative frequency. Variation in the order of head and modifier with respect to each other is indicated in column 8 in table 3.

Column 9 indicates whether the language involved uses morphological markers that identify a modifier within a predicate phrase, such as a particle used to mark manner modifiers. In this column, a + indicates that there is such a morphological marker, and a – that there isn't.

## 5. HYPOTHESES AND RESULTS

### 5.1 *General hypothesis*

After these preliminary introductions to our classification of parts-of-speech systems and of word order phenomena, we now turn to the relation between parts-of-speech systems and word order. The distinction among flexible, differentiated and rigid languages, central to our classification of parts-of-speech systems, provides the basis for the general hypothesis in (23).

- (23) The existence of a specialized lexical class in a language, i.e. a lexical class whose members are tied to one syntactic slot, makes it less necessary for this language to mark this slot and the phrase within which this slot occurs syntactically or morphologically; conversely, the existence of a flexible lexical class in a language, i.e. a lexical class whose members may occur in various syntactic slots, makes it more necessary for this language to mark these slots and the phrases within which these slots occur syntactically or morphologically.

The central idea behind this hypothesis is that lexical specialization is an important factor contributing to disambiguation in the processing of constituents. Where lexical specialization is absent, additional disambiguating

---

[15] When a language uses agreement with both attributively and predicatively used lexemes, agreement does not uniquely identify the modifier and is therefore not listed.

strategies will have to be invoked, i.e. there is a trade-off between lexical type on the one hand and syntactic and morphological structure on the other.

Ambiguity has been a central issue in psycholinguistic research since the 1960s, mainly in investigating to what extent language processing is guided by syntactic or semantic strategies.<sup>16</sup> Traditionally, two kinds of (temporary) ambiguity are distinguished: lexical and syntactic. An example from Clark & Clark (1977: 81) illustrates:

(24) *Lexical ambiguity*

After taking the **right** turn at the intersection, I ...

(25) *Syntactic ambiguity*

Knowing that **visiting relatives** could be tiresome, I ...

In the case of flexible lexemes, however, we are not dealing with lexical or syntactic ambiguity (recall that flexible lexemes are semantically vague, i.e. they are not polysemous; see section 3.7 above), but rather with FUNCTIONAL AMBIGUITY. When a flexible lexeme is being processed, the hearer has to determine how this lexeme is used, e.g. whether it serves as the head of a referential phrase (nominal function) or as a modifier of the head of the referential phrase (adjectival function).

In accordance with current psycholinguistic views on language processing (cf. Whitney 1998: 207, 222), we assume that

- (i) people utilize a strategy of ‘Immediacy of Interpretation’, which means that they try to interpret each word as completely as possible as soon as the word is received;
- (ii) sentence processing is guided by multiple kinds of information (syntactic, semantic, pragmatic, visual, prosodic, etc.);
- (iii) ambiguity obstructs rapid and efficient language processing; in such cases the hearer will look for contextual and other clues to direct sentence comprehension.

Since flexible lexemes are potentially ambiguous in that there is no lexically coded information as to the precise function of the lexeme in the actual linguistic expression, we hypothesize that languages with flexible lexemes will have certain morphosyntactic strategies at their disposal that provide the hearer with clues as to the correct interpretation of the flexible lexeme.

---

[16] More recently research on ambiguity has focused on the question whether language processing is modular or integrative (for an overview, see, for example, Whitney 1998: chapter 7). In the modular view, sentence comprehension occurs in a series of distinct and independent stages. According to the integrative view, on the other hand, sentence processing is guided by various subprocesses which are in close communication with each other. Recent psycholinguistic research has produced strong support for the integrative view (Whitney 1998: 231).

Morphosyntactically, there are basically two possibilities to guide the hearer as regards the intended function of a flexible lexeme in the actual linguistic expression: (i) morphological markers<sup>17</sup> in the immediate linguistic context; and (ii) syntactic patterning.<sup>18</sup> We take the second strategy as our point of departure, but we will come across languages in which morphological marking has the same disambiguating function as word order has in others. The nature of the potential functional ambiguity in flexible languages depends on the degree of flexibility of their parts-of-speech system. The absence of a verb/non-verb distinction leads to other potential functional ambiguities other than the absence of an adjective/adverb distinction. The following sections present more concrete partial hypotheses, starting with languages with the most flexible parts-of-speech systems (types 1 and 1/2) in section 5.2. Then, in section 5.3, we will include in our predictions languages with part-of-speech systems of types 1–2/3, and finally, in section 5.4, we will be concerned with all languages that have a flexible word class in their parts-of-speech systems (types 1–3/4).

## 5.2 *The verb/non-verb distinction*

### 5.2.1 *Basic word order at the clause level*

In languages without a distinct class of verbs, i.e. type 1 and 1/2 in table 2 above, lexical information is insufficient for the identification of the predicate phrase and the referential phrases within a sentence, given that there are no separate lexical classes that 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, we are here particularly concerned with the position of the main predicate. We predict that in these languages the main predicate should occupy a uniquely identifiable position under all circumstances. Since only initial and final positions in the sentence are uniquely identifiable,<sup>19</sup> languages of types 1 and 1/2 are not expected to have predicate-medial basic word order, unless the problem of identifying the constituents of the clause is solved by morphological means.

In table 4 (in which the data from table 3 are reshuffled in such a way that the languages are ordered on the basis of their parts-of-speech systems, from the most flexible to the most rigid), the upper part of column 1 shows that this expectation is borne out. In fact, the data show that the hypothesized

---

[17] Note that we use the term ‘morphological marker’ to refer to a segmental grammatical means of expression, be it a free grammatical word (particle) or a bound morpheme.

[18] A third disambiguating strategy is the use of prosody. Because of the lack of data we cannot systematically discuss this strategy in this paper.

[19] Even a rigid V2 position is not uniquely identifiable, since the number of words that make up the first constituent may vary.

## PARTS-OF-SPEECH SYSTEMS AND WORD ORDER

Language	PoS	Clause			Referential phrase			Predicate phrase		
		1 BWO	2 FixSP	3 Morph	4 BWO	5 FixSP	6 Morph	7 BWO	8 FixSP	9 Morph
Samoan	1	1	-	+	HM	+	-	HM	+	-
Tagalog	1	1	-	+	MHM	-	+	irr	irr	irr
Mundari	1/2	3	+	irr	MH	+	-	MH	+	-
Hurrian	2	3	+	irr	MH	+	-	MH	+	-
Quechua, Imbabura	2	3	-	+	MH	+	-	MH	+	-
Warao	2	3	+	irr	HM	+	-	MH	+	+
Turkish	2/3	3	+	irr	MH	+	-	MH	+	-
Ket	3	3	+	irr	MH	+	-	MH	+	-
Miao	3	2	+	irr	HM	+	-	HM	+	-
Ngiti	3	2/3	+	irr	MH	+	-	MHM	-	-
Tidore	3	2	+	irr	HM	+	-	HM	+	-
Lango	3/4	2	+	irr	HM	+	+	HM	+	+
Abkhaz	4	3	-	-	HM	-	-	MH	+	-
Arapesh	4	2	+	irr	MHM	-	-	MHM	-	-
Babungo	4	2	+	irr	HM	+	-	HM	+	-
Bambara	4	3	+	irr	HM	+	+	HM	+	-
Basque	4	3	-	-	HM	-	-	MH	+	-
Burushaski	4	3	-	-	MH	-	-	MH	-	-
Georgian	4	2/3	-	-	MH	+	-	MH	+	-
Hittite	4	3	+	irr	MHM	-	-	MH	+	-
Hungarian	4	2/3	-	-	MH	+	-	MH	+	-
Itelmen	4	2/3	+	irr	MHM	-	-	MH	+	-
Japanese	4	3	+	irr	MH	+	-	MH	+	-
Nama	4	3	-	-	MH	+	-	MH	+	-
Ngalakan	4	1/2	-	-	HM	-	-	MH	+	-
Polish	4	2	-	-	MHM	-	-	MHM	-	-
Koasati	4/5	3	-	-	HM	+	-	MHM	-	-
Nasioi	4/5	3	-	-	HM	-	-	MH	+	-
Paiwan	4/5	1	-	+	HM	-	+	HM	+	+
Pipil	4/5	1	-	-	MH	-	-	MH	-	-
Sumerian	4/5	3	+	irr	HM	-	-	MH	+	-
Alamblak	5	3	-	-	MHM	-	-	irr	irr	irr
Berbice Dutch	5	2	+	irr	MH	+	-	irr	irr	irr
Guarani	5	2	-	-	HM	+	-	irr	irr	irr
Kayardild	5	1/2/3	-	-	MH	-	-	irr	irr	irr
Kisi	5	2/3	+	irr	HM	+	+	irr	irr	irr
Oromo	5	3	+	irr	HM	+	-	irr	irr	irr
Wambon	5	3	+	irr	HM	+	-	irr	irr	irr
Mandarin Chinese	5/6	2	+	irr	MH	+	-	irr	irr	irr
Garo	5/6	3	+	irr	HM	+	-	irr	irr	irr
Gude	5/6	1/2	-	-	MH	+	-	irr	irr	irr
Nung	5/6	2	-	-	HM	+	-	irr	irr	irr
Tamil	5/6	3	-	-	MH	+	-	irr	irr	irr
West Greenlandic	5/6	3	-	-	HM	+	-	irr	irr	irr
Hixkaryana	6	2	-	-	irr	irr	irr	irr	irr	irr
Krongo	6	1/2	-	-	irr	irr	irr	irr	irr	irr
Navaho	6	3	+	irr	irr	irr	irr	irr	irr	irr
Nivkh	6	3	+	irr	irr	irr	irr	irr	irr	irr
Nunggubuyu	6	1/2/3	-	-	irr	irr	irr	irr	irr	irr
Tuscarora	6/7	2	-	-	irr	irr	irr	irr	irr	irr

Table 4  
Parts-of-speech systems and word order

correlation holds for a larger group of languages than predicted. Predicate-medial basic order is not only absent in languages of types 1 and 1/2 but also in languages of types 2 and 2/3.<sup>20</sup> We will present a possible explanation for this fact after discussing word order variation at the sentence level.

[20] In connection with the correlation observed it is worth noting that the two language families showing the largest number of languages with an extremely flexible parts-of-speech system, Polynesian and Salish, are consistently predicate-initial.

5.2.2 *Word order variation at the clause level*

A second way of looking at word order at the clause level in view of the absence of a verb/non-verb distinction concerns the question to what extent languages display deviant ordering patterns. The expectation is that flexible languages of types 1–1/2 will be more reluctant to allow such deviation, because any deviation from the basic pattern can lead to functional ambiguity as regards the identification of predicate and referential phrases. Since the number of referential phrases in argument position may differ, we consider here only the deviations in order between subject and predicate phrase.

The upper part of column 2 in table 4 shows us that two out of the three languages of type 1 and 1/2 actually do allow variation in the ordering of subject and predicate phrase. But column 3 shows that – as predicted by our general hypothesis (see section 5.1) – these languages solve the problem of the potential functional ambiguity arising from an alternative order by inserting special morphological markers in the non-basic ordering patterns, rather than by the ordering patterns themselves.

In Samoan, a predicate-initial language with a parts-of-speech system of type 1, placement of any referential phrase in sentence-initial position, before the predicate phrase, is accompanied by the addition of the presentative particle ‘*o*’, as illustrated in (27). The basic order is illustrated in (26).

*Samoan* (Mosel & Hovdhaugen 1992: 52, 56)

- (26) ‘Ua ō tamaiti i Apia.  
 PERF go children LD Apia  
 ‘The children have gone to Apia.’
- (27) ‘O le maile sã fasi e le teine.  
 PRES ART dog PAST hit ERG ART girl  
 ‘The dog was hit by the girl.’

In Tagalog, another predicate-initial language of type 1, it is the predicate rather than the preposed constituent that is obligatorily marked in cases of inversion. Example (29) shows the use of the predicate marker *ay*, which is absent when the predicate is in initial position, as in (28).

*Tagalog* (Himmelmann forthcoming b)

- (28) Ma-saráp ang pag-kain.  
 STAT-satisfaction SPEC GER-eating  
 ‘The food was good.’
- (29) Silá mag-iná ay na-ulog na.  
 3.PL RCP-mother PM REAL.STAT-sleep now  
 ‘The mother and her daughter fell asleep.’

We may contrast the situation obtaining in Samoan and Tagalog with the placement of the subject before the predicate phrase in Pipil, a



predicate-initial language with a parts-of-speech system of type 4/5. In this language, the subject may be emphasized by being placed in sentence-initial position. As a comparison of (30) and (31) shows, preposing of the subject does not require any additional morphological marking in Pipil.

*Pipil* (Campbell 1985: 103, 104)

(30)  $\emptyset$ -ki-miktih                    ne wa:kax ne ta:ka-t.  
 3.SG.SBJ-3.SG.OBJ-kill DEF COW        DEF man  
 'The man killed the cow.'

(31) Ne i-siwa:-w                     $\emptyset$ -ki-miktih.  
 ART 3.SG.POSS-wife-POSS 3.SG.SBJ-3.SG.OBJ-kill  
 ne chumpipi.  
 DEF turkey  
 'His wife killed the turkey.'

In the previous section we showed that our predictions concerning basic word order in languages of type 1 and 1/2 do in fact hold for languages up to and including type 2/3. The same seems to be true of our predictions concerning word order variation, now up to and including type 3/4. The upper part of column 2 in table 4 shows that the remaining languages of these types do not allow word order variation, with the exception of Imbabura Quechua. This language, however, applies a morphological disambiguating strategy too. In Imbabura Quechua, a predicate-final language with a parts-of-speech system of type 2, placement of the subject after the predicate phrase is accompanied by the obligatory addition of the topic marker *-ka*, as shown in (33). The basic order is shown in (32).

*Imbabura Quechua* (Cole 1982: 70, 71)

(32) Juan chagra-ta trabaja-ju-n.  
 Juan field-ACC work-PROGR-3  
 'Juan is working in the field.'

(33) Jatun wasi-ta chari-n Marya-ka/\*Marya.  
 big house-ACC have-3 Maria-TOP/Maria  
 'Maria has a big house.'

This may be compared to the placement of the subject after the predicate phrase in Basque, another predicate-final language, but one with a parts-of-speech system of type 4. Basic order in Basque is illustrated in (34). Example (35) shows that no additional morphological marking is needed when the subject is postposed.<sup>21</sup>

[21] Note that Basque is morphologically ergative but syntactically accusative (Saltarelli 1988: xiii).

*Basque* (Saltarelli 1988: 67)

- (34) Aita-k ama-ri gona gorri-a  
 father-SG.ERG mother-SG.DAT skirt red-SG.ABS  
 erosi d- $\emptyset$ -io- $\emptyset$ .  
 buy 3.ABS-PRS.AUX-3.SG.DAT-3.SG.ERG  
 'Father has bought a red skirt for mother.'
- (35) Gona gorria erosi d- $\emptyset$ -io- $\emptyset$   
 skirt red-SG.ABS buy 3.ABS-PRS.AUX-3.SG.DAT-3.SG.ERG  
 ama-ri aita-k.  
 mother-SG.DAT father-SG.ERG  
 'Father has bought a red skirt for mother.'

These examples reveal that flexible languages which do not use word order to unambiguously signal the position of the predicate phrase within the sentence use morphological means to fulfil the same disambiguating function. This is strong confirmation for our general hypothesis, which claims that the absence of lexical specialization requires syntactic or morphological marking of syntactic slots. In the presence of lexical specialization, these syntactic slots would have been identifiable on the basis of lexical information. The data in table 4, combined with the observations on the languages that at first sight seemed to be counterexamples, i.e. Samoan, Tagalog and Quechua, lead us to the generalization that languages with parts-of-speech systems 1-3/4 do not allow variation in the ordering of subject and predicate phrase, unless the deviation of the basic word order pattern is marked by morphological means. This is in sharp contrast with the situation obtaining in languages of types 4-6/7: 22 out of the 38 languages of these types do allow word order variation between subject and predicate without marking the deviation of the basic word order pattern by morphological means.

Note that our hypothesis does not predict the reverse, i.e. it does not exclude the existence of languages which do not have a flexible parts-of-speech system, but which do have a basic clause-initial or clause-final order and a fixed order of subject and predicate, or a morphological marker of a deviant order. For instance, Bambara, which has a parts-of-speech system of type 4, is predicate-final and does not allow alternations in the order of subject and predicate; Paiwan, which has a parts-of-speech system of type 4/5, is predicate-initial and uses a morphological marker when the order of subject and predicate deviates from the basic pattern.

It is noteworthy that, as in the case of basic word order at the clause level, our generalization with respect to word order variation extends over a wider range of languages (1-3/4) than predicted (1-1/2). An explanation of these facts is that we define our generalizations in terms of the position of the predicate, irrespective of the lexical class of the predicate. Earlier research (Hengeveld 1992b) has revealed, however, that the more flexible the parts-of-speech system of a language is, the higher the extent to which various classes

of lexemes may be used predicatively, often without the intervention of a copula. Consider in this respect the following Turkish examples:

*Turkish* (Lewis 1967)

- (36) Yol uzun.  
road long  
'The road is long.'
- (37) uzun yol  
long road  
'the long road'

Turkish has a parts-of-speech system of type 2/3. In Turkish, as in most languages of types 1–3/4, bare nominal stems may be used predicatively, which introduces a further potential functional ambiguity in the grammar as regards the identification of predicate phrases and referential phrases. Fixed word order patterns and/or morphological marking by means of copulas, for example, help to solve this problem of functional ambiguity.

### 5.3 *The noun/modifier distinction: word order variation within referential phrases*

In flexible languages without a separate class of nouns (types 1–2/3), the potential functional ambiguity arising from the nature of their parts-of-speech system concerns the interpretation of a lexical element as a head or a modifier within a referential phrase, since the interpretation of a non-verb as the head or a modifier of a referential phrase may interfere with its potential interpretation as the head or a modifier of the same or a contiguous referential phrase. We therefore predict that the order of head and modifier in referential phrases is fixed, unless there is a special morphological procedure which uniquely identifies the head-modifier relation within the referential phrase.

The upper parts of columns 4 and 5 in table 4 show that one out of the seven languages of types 1–2/3, Tagalog, actually does allow variation in the ordering of head and modifier within the referential phrase. However – as predicted by our general hypothesis – the problem of functional ambiguity is solved in this language by morphological means, as shown in column 6. Since head and modifier of referential phrases are linked to each other through a special morphological marker, it is always clear which elements go together within a referential phrase. Compare the following examples:

*Tagalog* (Himmelman forthcoming a: 5)

- (38) ulól na unggó  
foolish LNK monkey  
'foolish monkey'
- (39) unggó=ng ulól  
monkey=LNK foolish  
'foolish monkey'

Given that Tagalog is a language with a parts-of-speech system of type 1, and therefore does not make a distinction between nouns on the one hand and modifiers or adjectives on the other, it is hard to say whether example (38) should be glossed as ‘foolish monkey’ or as ‘monkey fool’, and the same goes for (39). But whatever the analysis, the fact that the linker *na/ng*<sup>22</sup> always occurs in between head and modifier ensures that the two are interpreted as being part of the same referential phrase.

It is worth noting that the generalization has a somewhat wider scope again: it applies to languages with a parts-of-speech system up to type 3/4. None of these languages allows word order freedom within the noun phrase without morphological marking. This may be contrasted with the situation in Itelmen, for example. In this language, with a parts-of-speech system of type 4, a modifier may occur on either side of the head, as in Tagalog. However, it may do so in the absence of additional morphological procedures, witness the following examples:

*Itelmen* (Georg & Volodin 1999: 108f.)

(40) qeʔm °plah  
mine big  
‘big mine’

(41) °plah massu  
big bear  
‘big bear’

In fact, while variation in the order of head and modifier within referential phrases without morphological marking is not found in languages of types 1–3/4, it is abundant in languages of types 4–5/6: in 13 of the 32 languages belonging to these types, the modifier does not have a fixed position in the absence of a morphological strategy. Furthermore, all 13 languages fall within types 4–5, i.e. they are languages with an open class of true adjectives. If we restrict ourselves to these types, then the proportion of languages rises to 50%: 13 out of the 26 languages of types 4–5 do show flexibility in the order of heads and modifiers within referential phrases, and do not use additional morphological strategies. We will return to this remarkable fact in section 6.4.

Note that, conversely, our hypothesis does not predict that there should not be languages that do not have a flexible parts-of-speech system, but which do use a morphological marker of attribution even in the absence of word order freedom within the referential phrase. In our sample, Lango, Bambara and Kisi are cases in point.

---

[22] The different forms of the linker are phonologically conditioned.

5.4 *The adjective/manner adverb distinction*5.4.1 *Basic word order in the predicate phrase*

In languages without a specialized class of adverbs (1-3/4), the interpretation of a lexical element as a modifier of the predicate phrase may interfere either with its potential interpretation as the head of that same predicate phrase or with its interpretation as a lexical modifier of a contiguous referential phrase. The latter issue will be discussed in section 5.4.2. Here, we first discuss potential functional ambiguity within the predicate phrase.

We observed in section 5.2.2 that the ordering possibilities in languages with parts-of-speech systems of types 1-3/4 are restricted by the fact that all classes of lexemes may be used predicatively. As a result, in this type of language, the lexical modifier of a predicate should preferably not occur in a position in which it might be interpreted as the head of the predicate phrase. Thus, we arrive at the following predictions:

- (i) In predicate-final languages with HM order in predicate phrases, the modifier of the predicate phrase appears in a position in which it might be interpreted as the main predicate. Since this leads to potential functional ambiguity in languages with parts-of-speech systems 1-3/4, HM order in predicate phrases is not expected to occur in these languages, unless the potential ambiguity is resolved by morphological means.
- (ii) In predicate-initial languages with MH order in predicate phrases, the modifier of the predicate phrase appears in a position in which it might be interpreted as the predicate. Since this leads to potential functional ambiguity in languages with parts-of-speech systems 1-3/4, MH order in predicate phrases is not expected to occur in these languages, unless the potential ambiguity is resolved by morphological means.
- (iii) In predicate-medial languages the modifier of the predicate phrase may always appear in a position in which it might be interpreted as the predicate, irrespective of the order of head and modifier in predicate phrases. HM order, MH order, or both orders may, therefore, be expected to occur in predicate-medial languages, including those with parts-of-speech systems 1-3/4. Columns 7 and 8 in table 4 show that this is correct for all languages of types 1-3/4 in the sample.

The problem of potential functional ambiguity in predicate-medial languages is generally resolved by morphological means other than the uniquely identifying ones that we have concentrated on so far. Thus, Lango inflects potential modifiers when used predicatively, Ngiti makes use of an obligatory copula, and Miao and Tidore use some tense/mood/aspect and/or person markers with potential modifiers used predicatively. Since in the case of the latter two languages these markers are not obligatorily present, these

languages seem to allow some functional ambiguity, as in the following example from Miao:

*Miao* (Harriehausen 1990: 147f.)

(42) daim tiab dawb

CL skirt white

‘the white skirt’ or ‘The skirt is white.’

It might be that in this case prosodic differences help disambiguate the two readings (see note 18).

#### 5.4.2 *Basic word order in referential phrases*

In flexible languages that do not make a distinction between adjectives and manner adverbs (types 1–3/4), a further potential functional ambiguity arising from their parts-of-speech system concerns the fact that the interpretation of a lexical element as a modifier within a referential phrase may interfere with its potential interpretation as a modifier within a predicate phrase. Since, as we showed in the previous section, the position of the modifier is, at least partly, determined by the fact that it may be mistaken for the predicate, the question now is what the consequences of this are for the position of the modifier within the referential phrase.

The difficulty of distinguishing between the use of a lexical element as a modifier of the referential phrase or of the predicate phrase in languages of types 1–3/4 leads to a number of predictions concerning basic word order at the level of the referential phrase. With respect to these orders, the following predictions may be formulated:

- (i) In predicate-final languages with HM order in referential phrases, the modifier of the referential phrase appears in a position contiguous to the modifier slot of the (MH) predicate phrase. Since this leads to potential functional ambiguity in languages with parts-of-speech systems 1–3/4, HM order in referential phrases is not expected to occur in these languages. This is represented schematically, and illustrated by means of pseudo-English examples, in (43).

#### *Predicate-final languages*

- (43) (a) ( $M_{\text{RefPhr}}$   $H_{\text{RefPhr}}$ ) ( $M_{\text{PrPhr}}$   $H_{\text{PrPhr}}$ )  
 (beautiful girl) (dance)  
 ‘The beautiful girl danced.’
- (b) \*( $H_{\text{RefPhr}}$   $M_{\text{RefPhr}}$ ) ( $M_{\text{PrPhr}}$   $H_{\text{PrPhr}}$ )  
 (girl beautiful) (dance)  
 ‘The beautiful girl danced.’  
 (girl) (beautiful dance)  
 ‘The girl danced beautifully.’

- (ii) Similarly, in predicate-initial languages with MH order in referential phrases, the modifier of the referential phrase appears in a position contiguous to the modifier slot of the (HM) predicate phrase. Since this, again, leads to potential functional ambiguity in languages with parts-of-speech systems I-3/4, MH order in referential phrases is not expected to occur in these languages. This is illustrated in (44).

*Predicate-initial languages*

- (44) (a) ( $H_{PrPhr} M_{PrPhr}$ ) ( $H_{RefPhr} M_{RefPhr}$ )  
 (dance ) (girl beautiful )  
 ‘The beautiful girl danced.’  
 (b) \*( $H_{PrPhr} M_{PrPhr}$ ) ( $M_{RefPhr} H_{RefPhr}$ )  
 (dance beautiful) ( girl )  
 ‘The girl danced beautifully.’  
 (dance ) (beautiful girl )  
 ‘The beautiful girl danced.’

- (iii) In predicate-medial languages, the modifier of the referential phrase appears in a position contiguous to the modifier slot when the order of head and modifier within referential phrases is the inverse of the order of head and modifier within predicate phrases. Since this, again, leads to potential functional ambiguity in languages with parts-of-speech systems I-3/4, the ordering of head and modifier within referential and predicate phrases is expected to be identical in these languages. This is illustrated in (45) and (46).

*Predicate-medial languages*

- (45) (a) ( $H_{RefPhr} M_{RefPhr}$ ) ( $H_{PrPhr} M_{PrPhr}$ ) ( $H_{RefPhr} M_{RefPhr}$ )  
 (girl beautiful) (sing ) (song nice )  
 ‘The beautiful girl sang a nice song.’  
 (b) \*( $M_{RefPhr} H_{RefPhr}$ ) ( $H_{PrPhr} M_{PrPhr}$ ) ( $M_{RefPhr} H_{RefPhr}$ )  
 (beautiful girl ) (sing nice ) ( song )  
 ‘The beautiful girl sang a song nicely.’  
 (beautiful girl ) (sing ) (nice song )  
 ‘The beautiful girl sang a nice song.’  
 (46) (a) ( $M_{RefPhr} H_{RefPhr}$ ) ( $M_{PrPhr} H_{PrPhr}$ ) ( $M_{RefPhr} H_{RefPhr}$ )  
 (beautiful girl ) ( sing ) (nice song )  
 ‘The beautiful girl sang a nice song.’  
 (b) \*( $H_{RefPhr} M_{RefPhr}$ ) ( $M_{PrPhr} H_{PrPhr}$ ) ( $H_{RefPhr} M_{RefPhr}$ )  
 (girl beautiful) ( sing ) (song nice )  
 ‘The beautiful girl sang a nice song.’  
 (girl ) (beautiful sing ) (song nice )  
 ‘The girl sang a nice song beautifully.’

Summarizing the predictions, we expect that within the group of languages of types I-3/4, MH order in referential phrases is strongly preferred in

predicate-final languages, HM order in predicate-initial languages, and consistent MH or HM order across referential and predicate phrases in predicate-medial languages. The data in the upper part of column 4 in table 4 appear to show that there are three counterexamples to these expectations, namely Tagalog, Warao and Ngiti.

Tagalog is not a real counterexample, since as stated in section 3.4, this language lacks a modifier slot within the predicate phrase, so that there is never potential functional ambiguity with the modifier of referential phrases.

Warao is a predicate-final language with a parts-of-speech system of type 2, which nevertheless has HM order in referential phrases,<sup>23</sup> as shown in (47) and (48).

*Warao* (Vaquero 1965: 143, 50)

(47) Arubuko ine obono-ya.  
bread I want-PRS  
'I want bread.'

(48) noboto sanuka  
child small  
'small child'

However, as indicated in column 9 in table 4, the problem of distinguishing between the interpretation of a lexical element as a modifier of the referential phrase or as a modifier of the predicate phrase is solved in Warao by the possibility of turning the lexical element in its manner reading into the head of a noun phrase provided with the postposition *tane* 'manner', thus resolving the problem of functional ambiguity raised by the ordering patterns. Compare (49) and (50).

*Warao* (Romero-Figeroa 1997: 119, 71)

(49) Oko kuana yaota-te arone **yakera** nahoro-te ...  
we hardness work-NPAST although beauty eat-NPAST  
'Although we work hard and eat well, ....'

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

If this morphological solution to the problem of solving functional ambiguity is taken into account, Warao is not an exception to our generalization.

Ngiti has both predicate-medial and predicate-final order. In this language, the manner constituent occupies a special sentence-initial or

[23] Interestingly, Warao is also the only language within the group of flexible languages in which the object precedes the subject.



sentence-final position, used for adjuncts and adpositional phrases in general. Thus the manner constituent does not necessarily form a single contiguous phrase with the main predicate, as in the other languages considered so far. These two special positions at the sentence margins offer sufficient possibilities for avoiding functional ambiguity, particularly since the order of head and lexical modifier within the referential phrase is fixed.

By contrast, as can be seen in table 4, in languages with parts-of-speech systems 4 and 4/5 – i.e. the remaining languages with lexical modifiers at both the referential phrase and the predicate phrase levels – the word-order combinations avoided in languages of types 1–3/4 are in fact more numerous than the other combinations. Within the large group of predicate-final languages with parts-of-speech systems 4 and 4/5, Abkhaz, Basque, Hittite, Koasati, Nasioi and Sumerian all have HM as their basic word order or as one of their basic word orders in referential phrases without any additional morphological marking, whereas only Burushaski, Japanese and Nama have MH basic word order in referential phrases. The only predicate-initial language with a parts-of-speech system of types 4–4/5, Pipil, uses MH order in referential phrases, which again is the order that would be avoided in a language with parts-of-speech systems 1–3/4. Two of the three predicate-medial languages with a parts-of-speech system of types 4–4/5, Arapesh and Polish, do not have a consistent identical ordering of heads and modifiers in referential and predicate phrases, whereas only one, Babungo, does.

All these facts are strong confirmation of our hypothesis that the ordering of heads and modifiers within referential and predicate phrases is strongly determined by the parts-of-speech systems of the languages involved.

### 5.5 *Summary*

The preceding sections have demonstrated that there is a clear connection between the parts-of-speech system of a language and restrictions on word order within that language. Our general hypothesis – that languages need syntactic and morphological means to mark syntactic slots in those cases in which disambiguation through lexical specialization is absent – is confirmed. Flexible languages exhibit severe restrictions on their word order possibilities. In those cases in which these restrictions seem to be violated, morphological marking fulfils the same disambiguating function. This suggests that lexical typology may fruitfully supplement the results achieved in syntactic and morphological typology. In the next section we will elaborate this point one step further by showing how our results throw new light on various findings in the earlier typological literature, concentrating again on word order.

## 6. EARLIER TYPOLOGICAL WORK

6.1 *Introduction*

Our findings supplement earlier typological work on word order in that they help to refine existing universals, disprove some earlier claims and lead to alternative explanations for linguistic phenomena. We will illustrate this by looking at the relation between verb–object and noun–adjective order, in section 6.2, the distribution of basic word orders, in section 6.3, and adjective doubling, in section 6.4.

6.2 *The relation between verb–object and noun–adjective order*

A number of the word order universals suggested by Greenberg (1966) subsequently gave rise to the notion of consistent ordering (also referred to as Natural Serialization and The Head Parameter), i.e. a preference for languages to display either HM or MH order across all head/modifier pairs. The basic diagnostic of the HM or MH status of a language was taken to be the location of the object relative to the verb. Although originally adjectives were assumed to pattern like other modifiers, favouring AN order in OV languages and NA order in VO, Dryer (1988: 191; 1992: 95) laid waste to this assumption by showing that there was no correlation between the location of the object relative to the verb and the adjective relative to the noun. Our data suggest that Dryer's refutation of the correlation in question may not be fully justified. While overall the location of the object relative to the verb is not a good predictor of the location of the lexical modifier relative to the head in referential phrases, there is a subset of languages for which a clear correlation between the two can be discerned. In languages with parts-of-speech systems 1–3/4, predicate-initial order strongly correlates with HM order and predicate-final order with MH in referential phrases with a lexical modifier. Significantly, predicate-medial languages are not involved in this correlation, which vindicates Greenberg's original universal formulated with respect to SOV and VSO languages and not OV as opposed to VO ones. Thus our parts-of speech typology rehabilitates a correlation, be it in a somewhat modified form.

6.3 *The distribution of basic word orders*

On the basis of a sample containing over a thousand languages, Tomlin (1986) found that languages were distributed over the six possible orderings of verb (V),<sup>24</sup> subject (S) and object (O) in the following way:

(51)	SOV	SVO	VSO	VOS	OVS	OSV
	44.78%	41.79%	9.20%	2.99%	1.24%	0%

[24] Note that we use the term 'verb' rather than 'predicate', the term which we would prefer, in accordance with the terminology used in the sources we are discussing here.

Tomlin argued that the difference in distribution reflects the interaction of three general ordering principles: (i) the Theme (read: subject) First Principle, (ii) Verb-Object Bonding (...OV... or ...VO...), and (iii) the Animated First Principle. Since SOV languages adhere to all three principles, they are most frequent. Conversely, OSV languages violate the three principles and are for that reason extremely rare (see e.g. Derbyshire & Pullum 1981) and – in Tomlin’s sample – unattested. Hawkins (1994: 331–339), in turn, attributes the above distribution of basic orders to relative ease of processing. The algorithm which he has devised for calculating processing ease, the Early Immediate Constituent recognition algorithm (EIC), identifies SOV as the most optimal order in terms of processing efficiency for an OV language, and SVO order as the most optimal for a VO language. Further, it predicts that the two orders should be grammaticalized, in the unmarked case, ‘at the expense of all other orders’ (Hawkins 1994: 337). As Tomlin’s data show, while this prediction fares rather well in relation to the remaining two OV orders, it is somewhat less successful in regard to the remaining two VO ones. Although languages with verb-initial order are undoubtedly less frequent than the ones with verb-final order, they definitely do occur, and in some areas of the world even with high frequency. What, then, is the motivation for the grammaticalization of either of the two verb-initial orders? Hawkins’ EIC has no explanation to offer in this connection. The EIC identifies VSO as the only viable alternative to SVO in terms of ease of processing. Nonetheless, as SVO is always more efficient, it remains a mystery why VSO should ever be selected. Our investigation suggests that one of the factors underlying this choice is the parts-of-speech system of a language.

We have shown that languages with highly flexible parts-of-speech systems do not have verb-medial order, which – in our sample – means either OVS (Hixkaryana) or SVO. Our explanation for the absence of such order in flexible parts-of-speech system languages is that in verb-medial languages the predicate is not uniquely identifiable in terms of its location: in intransitive clauses it is final, in transitive medial, and in pro-drop languages often initial. Consequently, word order cannot be employed to determine whether a given lexeme is or is not being used as a predicate. As this cannot be established on the basis of the lexeme itself, the burden of doing so would fall on morphological marking. However, subjects in SVO languages are rather unlikely to be morphologically marked. First of all, SVO languages lack nominal case marking more often than, for example, SOV languages (see e.g. Siewierska 1996 and the references therein). And secondly, in those SVO languages that do display case marking, it is generally the object rather than the subject which has an overt marker. Therefore, unless the predicate bears special marking, which of the lexemes is the S and which the predicate will not be immediately clear. Basic verb-initial order avoids these problems of identification; the predicate is uniquely identifiable in terms of its initial location.

In the light of the above, the question arises why all languages with a flexible parts-of-speech system are not predicate-initial. If predicate-initial order is such a successful strategy for identifying the function of a lexically underspecified lexeme, might we not expect it to be favoured over predicate-final order? The answer is no. As a predicate-identifying strategy, predicate-final order is just as good as predicate-initial. What is at issue is consistency. If the predicate is always clause-final, there is no reason why an unmarked clause-initial lexeme should be misinterpreted as a predicate. In transitive clauses, however, there is a potential danger of interpreting the object as the (intransitive) predicate. We would therefore expect predicate-final languages with a flexible parts-of-speech system to be particularly likely to display overt morphological case marking of the object or verbal agreement marking, or both. All the relevant languages in the sample do so.

It needs to be mentioned that the fact that predicate-initial and predicate-final orders are equally good identifying predicate strategies as far as flexible parts-of-speech system languages are concerned finds indirect support from Hawkins' (2002) more recent work on the relationship between dependency strength and linear precedence. Hawkins argues that the dependency relations between a dependent and an independent category may differ in strength, as a function of both the formal and the semantic properties that are assigned by the independent category to the dependent one. The strongest dependency relations involve filler-gap dependencies, as in *wh*-questions or zero-marked relative clauses. These he calls full addition dependencies. Slightly weaker are partial addition dependencies, such as those obtaining between a reflexive pronoun and its clausal antecedent. Weaker still are dependencies involving the reduction in the semantic or syntactic range of a dependent category by the independent one, called range reduction. Included among such dependencies are case marking and semantic role assignment, and – crucially for us – polysemy and ambiguity reduction in functional categories and parts of speech. According to Hawkins, strong dependencies favour placing the independent category before the dependent one. Weaker dependencies or more symmetrical dependencies allow for either order. In languages with a flexible parts-of-speech system, as in any other language, it is the predicate which determines the semantic role and (to a certain extent) the case marking of its arguments. However, the predicate is also dependent on the non-predicate lexemes for its identification as the predicate. Thus, flexible lexemes enter into symmetrical dependency relations (with respect to different properties) with each other, even more so than in non-flexible languages. As the dependencies go in both directions, and involve range reduction, they are weak, in the sense defined above. Therefore, there should be no preference for positioning the more independent category before the more dependent one. Given the overall cross-linguistic dominance of predicate-final orders, the fact that among the few

flexible languages there are predicate-initial ones as well as predicate-final ones bears out this prediction.

#### 6.4 Doubling

Doubling refers to the placement of lexical modifiers in referential phrases both before and after the head. Although doubling is widely attested, it has proved to be very difficult to formulate valid generalizations with respect to the circumstances in which it tends to occur. An early observation by Greenberg (1966) with respect to the doubling of adjectives is his Universal 19.

(52) *Universal 19*

When the general rule is that the descriptive adjective follows, there may be a minority of adjectives which usually precede, but when the general rule is that descriptive adjectives precede, there are no exceptions.

In his study on word order universals, Hawkins (1983) tried to explain doubling in terms of his Doubling Hypothesis, which says that doubling only occurs in adjacent subtypes as defined by his Prepositional and Postpositional Noun Modifier Hierarchy. For example, the Prepositional Noun Modifier Hierarchy (PrNMH) in (53) permits the co-occurrences in (54) (Hawkins 1983: 75f.):

(53)  $\text{Prep} \supset ((\text{NDem} \supset \text{NNum} \supset \text{NA}) \& (\text{NA} \supset \text{NG}) \& (\text{NG} \supset \text{NRel}))$

(54) (a)  $\text{Prep} \& \text{NDem} \& \text{NNum} \& \text{NA} \& \text{NG} \& \text{NRel}$

(b1)  $\text{Prep} \& \text{DemN} \& \text{NNum} \& \text{NA} \& \text{NG} \& \text{NRel}$

(b2)  $\text{Prep} \& \text{NDem} \& \text{NumN} \& \text{NA} \& \text{NG} \& \text{NRel}$

(c)  **$\text{Prep} \& \text{DemN} \& \text{NumN} \& \text{NA} \& \text{NG} \& \text{NRel}$**

(d)  **$\text{Prep} \& \text{DemN} \& \text{NumN} \& \text{AN} \& \text{NG} \& \text{NRel}$**

(e)  $\text{Prep} \& \text{DemN} \& \text{NumN} \& \text{AN} \& \text{GN} \& \text{NRel}$

(f)  $\text{Prep} \& \text{DemN} \& \text{NumN} \& \text{AN} \& \text{GN} \& \text{RelN}$

The PrNMH predicts that doubling of adjectives only occurs in languages of subtypes c. and d. in the co-occurrences listed in (54).<sup>25</sup> However, Hawkins (1983: 76f., 87) also encountered various exceptions to the Doubling Hypothesis in his sample, some of which he attributed to historical processes in that he claimed that certain occurrences of non-adjacent doubling are the result of ‘various idiosyncratic and language particular factors [which] can retard the loss of some word order’ (p. 77).

As we have shown (in section 5.3), doubling is more favoured in flexible languages in which the adjective is not a clearly differentiated part of speech, i.e. in languages of types 1–3/4. In these languages, doubling of lexical

[25] Note that type (54d) would be excluded by Greenberg’s Universal 19.

modifiers in referential phrases, i.e. the occurrence of lexical modifiers both before and after the head, is impossible. The only exception is Tagalog, for the reasons given in sections 3.4 and 5.4.2. Thus, we are now in a position to offer an alternative explanation: doubling is only allowed in those languages in which the adjective is a clearly differentiated part-of-speech, i.e. in languages with parts-of-speech systems 4–5/6. Table 4 shows clearly that this is the case.

## 7. CONCLUSION

In this article we have shown that the nature of the parts-of-speech system of a language imposes restrictions on the syntactic properties of that language, which can be explained in terms of their disambiguating function. We have also shown that in those cases in which languages display unpredicted syntactic properties, they apply morphological solutions with the same disambiguating effect. Our findings furthermore have allowed us to propose a number of improvements to existing work in word order typology. The results thus show that lexical typology is a necessary addition to existing work on syntactic and morphological typology.

## REFERENCES

### A. DESCRIPTIONS OF SAMPLE LANGUAGES

#### *Abkhaz*

Hewitt, Brian George (1979). *Abkhaz* (Lingua Descriptive Studies 2). Amsterdam: North-Holland.

Spruit, Arie (1986). *Abkhaz studies*. Ph.D. dissertation, University of Leiden.

#### *Alamblak*

Bruce, Les (1984). *The Alamblak language of Papua New Guinea (East Sepik)* (Pacific Linguistics C81). Canberra: The Australian National University.

#### *Arapesh, Mountain*

Conrad, Robert J. & Wogiga, Kepas (1991). *An outline of Bukiyip grammar* (Pacific Linguistics C113). Canberra: The Australian National University.

#### *Babungo*

Schaub, Willy (1985). *Babungo* (Croom Helm Descriptive Grammars). London: Croom Helm.

#### *Bambara*

Brauner, Siegmund (1974). *Lehrbuch des Bambara*. Leipzig: Enzyklopädie.

#### *Basque*

Saltarelli, Mario (1988). *Basque* (Croom Helm Descriptive Grammars). London: Croom Helm.

#### *Berbice Dutch Creole*

Kouwenberg, Silvia (1994). *A grammar of Berbice Dutch Creole* (Mouton Grammar Library 12). Berlin: Mouton de Gruyter.

#### *Burushaski, Hunza/Nager*

Berger, Hermann (1998). *Die Burushaski-Sprache von Hunza und Nager* (Neuindische Studien 13; 3 vols.). Wiesbaden: Harrassowitz.

*Garo*

Burling, Robbins (1961). *A Garo grammar* (Deccan College Monograph Series 25). Poona: Deccan College.

*Georgian*

Hewitt, Brian George (1995). *Georgian: a structural reference grammar* (London Oriental and African Language Library 2). Amsterdam: Benjamins.

Vogt, Hans (1971). *Grammaire de la langue géorgienne* (Institutet for Sammenlignende Kulturforskning B57). Oslo: Universitetsforlaget.

*Guarani*

Gregores, Emma & Suárez, Jorge A. (1967). *A description of colloquial Guarani*. The Hague: Mouton.

*Gude*

Hoskison, James Taylor (1983). *A grammar and dictionary of the Gude language*. Ph.D. dissertation, Ohio State University.

*Hittite*

Friedrich, Johannes (1974). *Hethitisches Elementarbuch. I: Kurzgefasste Grammatik* (3 vols.). Heidelberg: C. Winter.

Hout, Theo P. J. van den (1995). *Hettitisch voor beginners*. Archeologisch-Historisch Instituut der Universiteit van Amsterdam.

*Hixkaryana*

Derbyshire, Desmond C. (1979). *Hixkaryana* (Lingua Descriptive Studies 1). Amsterdam: North-Holland.

*Hungarian*

Behrens, Leila (1982). *Zur funktionalen Motivation der Wortstellung: Untersuchungen anhand des Ungarischen*. München: Veröffentlichungen des Finnisch-Ugrischen Seminars an der Universität München.

Kenesei, István, Vago, Robert M. & Fenyvesi, Anna (1998). *Hungarian* (Routledge Descriptive Grammars). London: Routledge.

*Hurrian*

Speiser, Ephraim A. (1941). *Introduction to Hurrian* (Annual of the American School of Oriental Research 20). New Haven, CA: American School of Oriental Research.

*Imbabura Quechua*

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

*Itelmen*

Georg, Stefan & Volodin, Alexander P. (1999). *Die itelmenische Sprache: Grammatik und Texte* (Tunguso Sibirica 5). Wiesbaden: Harrassowitz.

*Japanese*

Hinds, John (1986). *Japanese* (Croom Helm Descriptive Grammars). London: Croom Helm.

*Kayardild*

Evans, Nicholas D. (1995). *A grammar of Kayardild* (Moutin Grammar Library 15). Berlin: Mouton de Gruyter.

*Ket*

Werner, Heinrich (1997). *Die ketische Sprache* (Tunguso Sibirica 3). Wiesbaden: Harrassowitz.

*Kisi*

Childs, Tucker G. (1995). *A grammar of Kisi* (Mouton Grammar Library 16). Berlin: Mouton de Gruyter.

*Koasati*

Kimball, Geoffrey D. (1991). *Koasati grammar* (Studies in the Anthropology of North American Indians). Lincoln, NE: University of Nebraska Press.

*Krongo*

Reh, Mechthild (1985). *Die Krongo-Sprache (niino mó-dì). Beschreibung, Texte, Wörterverzeichnis* (Kölner Beiträge zur Afrikanistik 12). Berlin: Reimer.

*Lango*

Noonan, Michael P. (1992). *Lango* (Mouton Grammar Library 7). Berlin: Mouton de Gruyter.

*Mandarin Chinese*

Li, Charles N. & Thompson, Sandra A. (1981). *Mandarin Chinese*. Berkeley: University of California Press.

*Miao*

Harriehausen, Bettina (1990). *Hmong Njua: syntaktische Analyse einer gesprochenen Sprache mithilfe datenverarbeitungstechnischer Mittel und sprachvergleichende Beschreibung des südostasiatischen Sprachraumes* (Linguistische Arbeiten 245). Tübingen: Niemeyer.

*Mundari*

Cook, Walter A. (1965). *A descriptive analysis of Mundari*. Ph.D. dissertation, Georgetown University.

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

*Nama Hottentot*

Hagman, Roy S. (1974). *Nama Hottentot grammar*. Ph.D. dissertation, Columbia University.

Rust, Friedrich (1965). *Praktische Namagrammatik* (Communications from the School of African Studies, University of Cape Town 31). Cape Town: Balkema.

*Nasioi*

Rausch, J. (1912). Die Sprache von Südost-Bougainville, Deutsche Salomonsinseln. *Anthropos* 7. 105–134. & 585–616. & 964–994.

*Navaho*

Young, Robert W. & Morgan, William Sr. (1987). *The Navajo language: a grammar and colloquial dictionary* (revised edn.). Albuquerque: University of New Mexico Press.

*Ngalakan*

Merlan, Francesca (1983). *Ngalakan grammar, texts, and vocabulary* (Pacific Linguistics B89). Canberra: The Australian National University.

*Ngiti*

Kutsch Lojenga, Constance (1993). *Ngiti* (Nilo-Saharan 9). Köln: Reimer.

*Nivkh*

Gruzdeva, Ekaterina (1998). *Nivkh* (Languages of the World: Materials 111). München: Lincom.  
Mattissen, Johanna & Drossard, Werner (1998). *Lexical and syntactic categories in Nivkh* (Theorie des Lexikons: Arbeiten des Sonderforschungsbereich 282, 85). Düsseldorf: Heinrich Heine Universität.

*Nung*

Saul, Janice E. & Freiburger Wilson, Nancy (1980). *Nung Grammar* (Summer Institute of Linguistics Publications in Linguistics 62). Dallas: Summer Institute of Linguistics & Arlington: University of Texas.

*Nunggubuyu*

Heath, Jeffrey (1984). *Functional grammar of Nunggubuyu* (AIAS New Series 53). Canberra: Australian Institute of Aboriginal Studies.



*Oromo, Boraana*

Stroomer, Harry J. (1995). *A grammar of Boraana Oromo (Kenya)* (Cushitic Language Studies **11**). Köln: Köppe.

*Paiwan*

Egli, Hans (1990). *Paiwangrammatik*. Wiesbaden: Harrassowitz.

*Pipil*

Campbell, Lyle (1985). *The Pipil language of El Salvador* (Mouton Grammar Library **1**). Berlin: Mouton de Gruyter.

*Polish*

Teslar, Joseph Andrew (1953). *A new Polish grammar* (6th edn.). Edinburgh: Oliver and Boyd.

*Samoan*

Mosel, Ulrike & Hovdhaugen, Even (1992). *Samoan reference grammar* (Instituttet for sammenlignende kulturforskning **B85**). Oslo: Scandinavian University Press.

*Sumerian*

Thomsen, Marie-Louise (1984). *The Sumerian language* (Mesopotamia: Copenhagen Studies in Assyriology **10**). Copenhagen: Akademisk Forlag.

*Tagalog*

Himmelman, Nikolaus (forthcoming a). Tagalog. In Booij, Geert, Lehmann, Christian & Mugdan, Joachim (eds.), *Morphology: a handbook on inflection and word formation* (part II). Berlin: Mouton de Gruyter.

Himmelman, Nikolaus (forthcoming b). Lexical categories and voice in Tagalog. In Austin, Peter & Musgrave, Simon (eds.), *Grammatical relations and voice in Austronesian*. Stanford: CSLI.  
Schachter, Paul & Otnes, Fe T. (1972). *Tagalog reference grammar*. Berkeley: University of California Press.

*Tamil*

Asher, Ron E. (1982). *Tamil* (Lingua Descriptive Studies **7**). Amsterdam: North-Holland.

*Tidore*

van Staden, Miriam (2000). *Tidore: a linguistic description of a language of the North Moluccas*. Ph.D. dissertation, Leiden University.

*Turkish*

Kornfilt, Jaklin (1997). *Turkish* (Routledge Descriptive Grammars). London: Routledge.

Lewis, Geoffrey L. (1967). *Turkish grammar*. Oxford: Oxford University Press.

*Tuscarora*

Mithun Williams, Marianne (1976). *A grammar of Tuscarora* (Garland Studies in American Indian Linguistics). New York: Garland.

*Wambon*

de Vries, Lourens & de Vries-Wiersma, Robinia (1992). *The morphology of Wambon* (Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde **151**). Leiden: KITLV Press.

*Warao*

Romero-Figeroa, Andres (1997). *A reference grammar of Warao* (Lincom Studies in Native American Linguistics **6**). München: Lincom.

Vaquero, P. Antonio (1965). *Idioma Warao: morfología, sintaxis, literotura*. Caracas: Estudio Venezolanos Indiganis.

*West Greenlandic*

Fortescue, Michael (1984). *West Greenlandic* (Croom Helm Descriptive Grammars). London: Croom Helm.

## B. IN-TEXT REFERENCES

- Behrens, Leila (1982). *Zur funktionalen Motivation der Wortstellung: Untersuchungen anhand des Ungarischen*. München: Veröffentlichungen des Finnisch-Ugrischen Seminars an der Universität München.
- Bhat, Darbhe Narayana Shankara (1994). *The adjectival category: criteria for differentiation and identification* (Studies in Language Companion Series 24). Amsterdam: Benjamins.
- Bisang, Walter. (1996). Areal typology and grammaticalization: processes of grammaticalization based on nouns and verbs in East and Mainland South East Asian languages. *Studies in Language* 20, 519–597.
- Burling, Robbins (1961). *A Garo grammar* (Deccan College Monograph Series 25). Poona: Deccan College.
- Campbell, Lyle (1985). *The Pipil language of El Salvador* (Mouton Grammar Library 1). Berlin: Mouton de Gruyter.
- Chomsky, Noam (1995). *The Minimalist program* (Current Studies in Linguistics 28). Cambridge, MA: MIT Press.
- Clark, Eve V. & Clark, Herbert H. (1977). *Psychology and language: an introduction to psycholinguistics*. New York: Harcourt Brace Jovanovich.
- Cole, Peter (1982). *Imbabura Quechua* (Lingua Descriptive Studies 5). Amsterdam: North-Holland.
- Croft, William (2001). *Radical construction grammar: syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Cruse, D. Alan (1986). *Lexical semantics*. Cambridge: Cambridge University Press.
- Derbyshire, Desmond & Pullum, Geoffrey K. (1981). Object initial languages. *International Journal of American Linguistics* 47, 192–214.
- de Vries, Lourens & de Vries-Wiersma, Robinia (1992). *The morphology of Wambon* (Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 151). Leiden: KITLV Press.
- Dixon, Robert M. W. (1980). *The languages of Australia*. Cambridge: Cambridge University Press.
- Dryer, Matthew S. (1988). Object-verb order and adjective-noun order: dispelling a myth. *Lingua* 74, 185–217.
- Dryer, Matthew S. (1992). The Greenbergian word order correlations. *Language* 68, 81–138.
- Dryer, Matthew S. (1997). On the six-way word order typology. *Studies in Language* 21, 69–103.
- Evans, Nicholas (2000). Word classes in the world's languages. In Booij, Geert, Lehmann, Christian & Mugdan, Joachim (eds.), *Morphology: a handbook on inflection and word formation* (part I). Berlin: Walter de Gruyter. 708–732.
- Georg, Stefan & Volodin, Alexander P. (1999). *Die itelmenische Sprache: Grammatik und Texte* (Tunguso Sibirica 5). Wiesbaden: Harrassowitz.
- Greenberg, Joseph H. (1966). Some universals of grammar with particular reference to the order of meaningful elements. In Greenberg, Joseph H. (ed.), *Universals of language* (2nd edn.). Cambridge, MA: MIT Press. 73–113.
- Harriehausen, Bettina (1990). *Hmong Njua: syntaktische Analyse einer gesprochenen Sprache mithilfe datenverarbeitungstechnischer Mittel und sprachvergleichende Beschreibung des südostasiatischen Sprachraumes* (Linguistische Arbeiten 245). Tübingen: Niemeyer.
- Hawkins, John A. (1983). *Word order universals: quantitative analyses of linguistic structure*. New York: Academic Press.
- Hawkins, John A. (1994). *A performance theory of order and constituency*. Cambridge: Cambridge University Press.
- Hawkins, John A. (2002). Symmetries and asymmetries: their grammar, typology and parsing. *Theoretical Linguistics* 28, 95–149.
- Hawkins, John A. & Gilligan, Gary (1988). Prefixing and suffixing universals in relation to basic word order. *Lingua* 74, 219–259.
- Hengeveld, Kees (1992a). Parts of speech. In Fortescue, Michael, Harder, Peter & Kristoffersen, Lars (eds.), *Layered structure and reference in a functional perspective* (Pragmatics and Beyond New Series 23). Amsterdam: Benjamins. 29–55.

- Hengeveld, Kees (1992b). *Non-verbal predication: theory, typology, diachrony* (Functional Grammar Series 15). Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus (forthcoming a). Tagalog. In Booij, Geert, Lehmann, Christian & Mugdan, Joachim (eds.), *Morphology: a handbook on inflection and word formation* (part II). Berlin: Mouton de Gruyter.
- Himmelmann, Nikolaus (forthcoming b). Lexical categories and voice in Tagalog. In Austin, Peter & Musgrave, Simon (eds.), *Grammatical relations and voice in Austronesian*. Stanford: CSLI.
- Hoffmann, John (1903). *Mundari grammar*. Calcutta: The Secretariat Press.
- Kutsch Lojenga, Constance (1993). *Ngiti* (Nilo-Saharan 9). Köln: Reimer.
- Lehmann, Christian (1990). Towards lexical typology. In Croft, William, Denning, Keith & Kemmer, Suzanne (eds.), *Studies in typology and diachrony: papers presented to Joseph H. Greenberg on his 75th birthday* (Typological Studies in Language 20). Amsterdam: Benjamins. 161–185.
- Lewis, Geoffrey L. (1967). *Turkish grammar*. Oxford: Oxford University Press.
- McGregor, William (2002). *Verb classification in Australian languages* (Empirical Approaches to Language Typology 25). Berlin: Mouton de Gruyter.
- Mithun Williams, Marianne (1976). *A grammar of Tuscaraora* (Garland Studies in American Indian Linguistics). New York: Garland.
- Mithun, Marianne (2000). Nouns and verbs in Iroquoian languages: multicategorisation from multiple criteria. In Vogel & Comrie (eds.), 397–420.
- Mosel, Ulrike & Hovdhaugen, Even (1992). *Samoan reference grammar* (Instituttet for sammenlignende kulturforskning B85). Oslo: Scandinavian University Press.
- Plank, Frans (1998). The co-variation of phonology with morphology and syntax: a hopeful history. *Linguistic Typology* 2. 195–230.
- Rijkhoff, Jan (2002). *The noun phrase* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.
- Rijkhoff, Jan (2003). When can a language have nouns and verbs? *Acta Linguistica Hafniensia* 35. 7–38.
- Rijkhoff, Jan, Bakker, Dik, Hengeveld, Kees & Kahrel, Peter (1993). A method of language sampling. *Studies in Language* 17. 169–203.
- Romero-Figeroa, Andres (1997). *A reference grammar of Warao* (Lincom Studies in Native American Linguistics 6). München: Lincom.
- Ruhlen, Merritt (1991). *A guide to the world's languages*, vol. 1: *Classification* (2nd edn.). Stanford: Stanford University Press. [1st edn. 1987.]
- Saeed, John I. (2003). *Semantics* (2nd edn.). Oxford: Blackwell.
- Saltarelli, Mario (1988). *Basque* (Croom Helm Descriptive Grammars). London: Croom Helm.
- Sasse, Hans-Jürgen (1988). *Der irokesische Sprachtyp* (Arbeitspapiere des Instituts für Sprachwissenschaft zu Köln, Neue Folge 9).
- Schachter, Paul (1985). Parts-of-speech systems. In Shopen, Timothy (ed.), *Language typology and syntactic description*, vol. I: *Clause structure*. Cambridge: Cambridge University Press. 3–61.
- Siewierska, Anna (1996). Word order type and alignment type. *STUF (Sprachtypologie und Universalienforschung)* 49. 149–176.
- Siewierska, Anna (1998). Variation in major constituent order: a global and a European perspective. In Siewierska, Anna (ed.), *Constituent order in the languages of Europe*. Berlin: Mouton de Gruyter. 475–551.
- Stassen, Leon (1997). *Intransitive predication* (Oxford Studies in Typology and Linguistic Theory). Oxford: Oxford University Press.
- Steele, Susan (1978). Word order variation: a typological survey. In Greenberg, Joseph H., Ferguson, Charles A. & Moravcsik, Edith A. (eds.), *Universals of human language*, vol. 4: *Syntax*. Stanford: Stanford University Press. 585–623.
- Tomlin, Russell S. (1986). *Basic word order: functional principles*. London: Croom Helm.
- Vaquero, P. Antonio (1965). *Idioma Warao: morfología, sintaxis, literatura*. Caracas: Estudio Venezolanos Indigianis.
- Vogel, Petra M. (2000). Grammaticalisation and part-of-speech systems. In Vogel & Comrie (eds.), 259–284.

- Vogel, Petra M. & Comrie, Bernard (eds.) (2000). *Approaches to the typology of word classes* (Empirical Approaches to Language Typology 23). Berlin: Mouton de Gruyter.
- Vogt, Hans (1971). *Grammaire de la langue géorgienne* (Institutet for Sammenlignende Kulturforskning B57). Oslo: Universitetsforlaget.
- Wetzer, Harrie (1996). *The typology of adjectival predication* (Empirical Approaches to Language Typology 17). Berlin: Mouton de Gruyter.
- Whitney, Paul (1998). *The psychology of language*. Boston & New York: Houghton Mifflin.
- Wilkins, David P. (2000). Ants, ancestors and medicine: a semantic and pragmatic account of classifier constructions in Arrernte (Central Australia). In Senft, Gunter (ed.), *Systems of nominal classification* (Language, Culture and Cognition 4). Cambridge: Cambridge University Press. 147–216.

*Authors' addresses: (Hengeveld)*

*Department of Theoretical Linguistics, University of Amsterdam,  
Spuistraat 210, NL-1012 VT Amsterdam, The Netherlands.  
E-mail: p.c.hengeveld@uva.nl*

*(Rijkhoff)*

*Department of Linguistics, University of Aarhus,  
Jens Chr. Skous Vej 7 (467–517), DK-8000 Aarhus C, Denmark.  
E-mail: linjr@hum.au.dk*

*(Siewierska)*

*Department of Linguistics and Modern English Language,  
Lancaster University, Lancaster LA1 4YT, U.K.  
E-mail: a.siewierska@lancaster.ac.uk*