2

Parts-of-speech systems as a basic typological determinant

KEES HENGEVELD

2.1 Introduction

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

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

After briefly summarizing the parts-of-speech theory proposed in Hengeveld (1992a, b) and refined in Hengeveld et al. (2004) in Section 2.2, I introduce, in Section 2.3, the predictions that follow from it with respect to other properties of

the languages involved. The predictions are presented in four groups, which have to do with the identifiability of constituents (Section 2.4), the formal integrity of lexemes (Section 2.5), the morphological and semantic unity of classes of lexemes (Section 2.6), and the pervasiveness of flexibility or rigidity in the grammar as a whole (Section 2.7). Section 2.8 then brings together the various results and gives a general characterization of types of languages in terms of combinations of lexical, morphological, and syntactic features.

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

Hengeveld (1992a, b) classifies basic and derived lexemes in terms of their distribution across the four functional slots given in Figure 2.1.

	HEAD	MODIFIER
PREDICATE PHRASE	verb	manner adverb
REFERENTIAL PHRASE	noun	adjective

Figure 2.1 Lexemes and functions

Figure 2.1 shows that the four functional positions are based on two parameters, one involving the opposition between predication and reference, the other between heads and modifiers. Together, these two parameters define the following four functions: (i) head of a predicate phrase, (ii) modifier of the head of a predicate phrase, (iii) head of a referential phrase, and (iv) modifier of the head of a referential phrase. The four functions and their lexical expression can be illustrated by means of the English sentence in (1).

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

English can be said to display separate lexeme classes of verbs, nouns, adjectives, and (derived) manner adverbs, on the basis of the distribution of these classes across the four functions identified in Figure 2.1: verbs like *sing* are used as heads of predicate phrases; nouns like *girl* as heads of referential phrases; adjectives like *tall* as modifiers of heads of referential phrases; and manner adverbs like *beautifully* as modifiers of heads of predicate phrases. Thus, in this example there is a one-to-one relation between function and lexeme class. Parts-of-speech systems of this type are called differentiated, as for each function there is a separate class of lexemes.

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

The four categories of lexemes in Figure 2.1 may be defined as follows: a verb (V) is a lexeme that can be used as the head of a predicate phrase only; a noun (N) is a lexeme that can be used as the head of a referential phrase; an adjective (A) is a lexeme that can be used as a modifier within a referential phrase; and a manner adverb (MAdv) is a lexeme that can be used as a modifier within a predicate phrase. Note that within the class of adverbs I restrict myself to manner adverbs. I exclude other classes of adverbs, such as temporal and spatial ones, since these do not modify the head of the predicate phrase, but rather modify the sentence as a whole. The restriction imposed on verbs that they can be used predicatively *only* is not paralleled in the definitions of the other lexeme classes, as these in many languages allow a predicative use apart from their distinguishing non-predicative functions.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

In sum, the difference between English (differentiated), Turkish (flexible), and Krongo (rigid) is thus that (i) Turkish has a class of flexible lexical items that may be used in several functions, where English uses three specialized classes (nouns, adjectives, and manner adverbs), and that (ii) Krongo lacks classes of lexical items for the modifier functions, where English does have lexical classes of adjectives and manner adverbs. Krongo has to resort to alternative syntactic strategies to compensate for the absence of a lexical solution. These differences may be represented as in Table 2.1

Table 2.1 Flexible, differentiated, and rigid languages

Language	*Head of predicate phrase*	*Head of referential phrase*	*Modifier of head of referential phrase*	*Modifier of head of predicate phrase*
Turkish	verb	non-verb	non-verb	non-verb
English	verb	noun	adjective	manner adverb
Krongo	verb	noun	–	–

As Table 2.1 shows, Turkish and Krongo are similar in that they have two main classes of lexemes. They are radically different, however, in the extent to which one of these classes may be used in the construction of predications: the Turkish class of non-verbs may be used in three functions, while the Krongo class of nouns may be used as the head of a referential phrase only. Notice that for a lexeme class to be classified as flexible, the flexibility should not be a property of a subset of items, but a general feature of the entire class.

Hengeveld (1992a, b) and Hengeveld et al. (2004) argue that the arrangement of the functions in Table 2.1 is not a coincidence. It is claimed to reflect the parts-of-speech hierarchy in (10):

(10) Head of pred. phrase > Head of ref. phrase > Modifier of head of ref. phrase > Modifier of head of pred. phrase

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

The hierarchy in (10), combined with the distinction between flexible, differentiated, and rigid languages, predicts a set of seven possible parts-of-speech systems, which is represented in Figure 2.2. As this figure shows, it is predicted that languages can display three different degrees of flexibility (systems 1–3), three different degrees of rigidity (systems 5–7), or can be differentiated (type 4). Of the languages discussed earlier, Turkish would be a type 2 language, English a type 4 language, and Krongo a

type 6 language. Note that I use the term 'contentive' for lexical elements that may appear in any of the four functions distinguished. The term 'modifier' is used for lexemes that may be used as modifiers in both predicative and referential phrases.

PoS-system		*Head of predicate phrase*	*Head of referential phrase*	*Modifier of head of referential phrase*	*Modifier of head 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 2.2 Parts-of-speech systems

In addition to the seven types listed in Figure 2.2, there are so-called intermediate systems, showing characteristics of two systems that are contiguous in the figure. In flexible languages the most common source for such an intermediate status is that derived stems show a lower degree of flexibility than basic stems. In rigid languages the most common source of an intermediate status is the existence of small, closed classes of lexemes at the fringe of the system. Figure 2.3 (see also Smit 2007) shows the full set of possible systems, including the intermediate ones.

An important point to be made is that the classification in Figure 2.3 is based on the properties of lexeme classes, not of word classes. Flexible lexemes, when put to use in a specific function, may receive inflections that are specific to that function. Thus, in the Turkish example (2) the lexeme *güzel* 'beauty', used as the head of a referential phrase, receives the possessive marker *-im* '1.POSS'. This possibility is lacking when the same lexeme is used as a modifier. The word *güzelim* 'my beauty' can thus be said to be a nominal word, but it is based on a lexeme that can be used flexibly in three different functions, each allowing different inflectional possibilities.

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

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

PoS-system		*Head of predicate phrase*	*Head of referential phrase*	*Modifier of head of referential phrase*	*Modifier of head of predicate phrase*
Flexible	1	contentive			
	1/2	contentive			
			non-verb		
	2	verb	non-verb		
	2/3	verb	non-verb		
				modifier	
	3	verb	noun	modifier	
	3/4	verb	noun	modifier	
					manner adverb
Differentiated	4	verb	noun	adjective	manner adverb
Rigid	4/5	verb	noun	adjective	(manner adverb)
	5	verb	noun	adjective	
	5/6	verb	noun	(adjective)	
	6	verb	noun		
	6/7	verb	(noun)		
	7	verb			

Figure 2.3 Parts-of-speech systems, including intermediate ones

2.3 Four sets of predictions

The approach to PoS-systems outlined in Section 2.2 leads to a number of predictions concerning the PoS-system of a language and other aspects of the grammar of that language. These predictions can be grouped together under four headings.

- Identifiability—The more specialized a lexical class is, i.e. the more it is tied to one functional slot, the less it is necessary to mark this slot and the phrase it forms part of syntactically or morphologically, i.e. there is a trade-off between lexical structure on the one hand and syntactic and morphological structure on the other.

An example of a prediction that follows from this observation is that rigid languages may be expected to display more freedom of word order than flexible languages.

- Integrity—The formal integrity of a lexeme, i.e. its formal independence of morphological material specific to a certain function, increases its applicability in various functions. An example of a prediction that follows from this observation is that flexible lexemes may be expected not to show morphologically conditioned stem alternation.
- Unity—The phonological, morphological, and semantic unity of a lexical class increases its applicability in various syntactic slots. An example of a prediction that follows from this observation is that intrinsic gender and conjugation classes may be expected not to occur in flexible languages.
- Pervasiveness—Flexibility and rigidity of lexical stems may be expected to correlate with functionality and rigidity of other morphological and syntactic units within the grammar and with functions not covered by the PoS-hierarchy. An example of a prediction that follows from this observation is that case-marked noun phrases or adpositional phrases may be expected to be used predicatively more readily in flexible languages than in rigid languages.

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

2.4 Identifiability

2.4.1 *Introduction*

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

2.4.2 *Clause*

In languages that do not have a distinct class of verbs, i.e. types 1 and 1/2 in Figure 2.3, lexical information is insufficient to arrive at the identification of the predicate phrase and the referential phrases within a sentence, given that there are no separate lexical

classes the members of which are used to fill the head slots of predicate phrases and referential phrases. Since the number of referential phrases in argument function in a sentence may vary, it is particularly the position of the main predicate that may help to disambiguate between the two types of phrase. Hengeveld et al. (2004) therefore predict that in these languages the main predicate should occupy a uniquely identifiable position under all circumstances. Since only an initial and a final position in the sentence are uniquely identifiable, they predict languages of types 1 and 1/2 not to have predicate medial basic word order, unless the problem of identifying the constituents of the clause is solved by segmental means. This prediction is confirmed. The following examples from Samoan (Mosel and Hovdhaugen 1992: 52, 56) illustrate the phenomenon:

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

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

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

In the sample used by Hengeveld et al. (2004) this also holds for languages of types 2 and 2/3. The explanation for this is that languages of these types allow all kinds of non-verbal constituents to be used predicatively (see Hengeveld 1992b). This again leads to further potential ambiguity as regards the interpretation of a constituent as a predicate phrase or a referential phrase, which can be solved by the same means as those listed above: rigid order and/or segmental marking. This is illustrated by the following examples from Turkish (Lewis 1967/1985):

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

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

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

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

2.4.3 *Phrase*

In all languages with some degree of flexibility, i.e. types 1 through 3/4 in Figure 2.3, there is potential ambiguity as regards the identification of heads and modifiers within and across predicate phrases and referential phrases. For instance, if a language has a class of flexible non-verbs, and a speaker uses these to fill the head and modifier slots of a referential phrase, lexical information is insufficient to decide which one is the head and which one the modifier; and if a language has a class of flexible modifiers rather than separate classes of adjectives and manner adverbs, freedom of constituent order creates a situation in which an addressee does not know whether to interpret a lexeme as the modifier of e.g. a preceding noun or of a following verb.

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

The prediction is borne out by the data: languages of types 1 through 3/4 have a fixed order of head and modifier or mark a deviation from this pattern segmentally[3], while languages of other types may or may not show such restrictions, and actually

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

often don't. A case of a language not respecting the word order restriction but repairing this morphologically is Warao. Consider the following examples (Vaquero 1965: 50; Romero-Figueroa 1997: 71):

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

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

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

2.4.4 *Summary of correlations*

The various properties of flexible languages that follow from the fact that constituents cannot be identified sufficiently on the basis of information that is intrinsic to the lexemes that are being used are summarized in Figure 2.4. The top row lists the PoS-systems in order of increasing rigidity, the blank boxes in between numbers representing the intermediate types.

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

Figure 2.4 Identifiability and PoS-system

2.5 Integrity

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

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

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

(17) a. *hûŋ*
come.HORT

b. *hûŋ* *lé*
come.HORT NEG

(18) a. *baa*
hang.HORT

b. *bee*
hang.HORT.NEG

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

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

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

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

As shown in Hengeveld (2007), the languages of his sample confirm all three testable predictions, that is, no flexible stem in any of the languages studied exhibits fusional morphology. Flexible stems only participate in the morphological processes of agglutination and isolation. An interesting consequence of this conclusion is that it

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

is not languages that should be classified in terms of their morphological type, but stem classes within languages.

The properties of flexible languages that follow from the fact that flexible stems need to be formally independent can be summarized as in Figure 2.5.[5]

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

Figure 2.5 Integrity and PoS-system

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

2.6 Unity

2.6.1 *Introduction*

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

2.6.2 *Morphological subclasses*

The morphological unity of a lexical class, i.e. the absence of intrinsic morphological subclasses (as opposed to semantic and phonological subclasses; see Corbett 1991) triggering specific morphological processes, increases its applicability in various functional slots. Taking this perspective, Hengeveld and Valstar (2010) hypothesize

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

that this type of differentiation within lexeme classes will be absent in flexible languages. More specifically, they hypothesize that:[6]

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

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

2.6.3 *Semantic subclasses*

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

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

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

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

Rijkhoff (2003) shows that in fact languages without a distinct class of verbs (i.e. types 1 and 1/2) never exhibit differentiation according to transitivity within their flexible lexeme classes, while languages with a distinct class of verbs always exhibit

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

this differentiation. This confirms the suggestion in Section 2.6.1 that the functional elasticity of a lexeme class does not combine very well with internal semantic differentiation within that class.

Another example of this is provided in Rijkhoff (2000) (see also Rijkhoff 2004), in which it is shown that in flexible languages without a differentiated class of nouns (i.e. types 1 through 2/3), the flexible lexemes are always transnumeral, i.e. not intrinsically specified for singular (or plural) number. A morphosyntactic reflex of this is that, when containing a numeral, the referential phrase is not simultaneously marked for number.[7] Consider the following examples from Turkish (Lewis 1967/1985: 26):

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

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

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

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

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

2.6.4 *Correlations*

The properties of flexible languages that follow from the fact that flexible stems need to be undifferentiated both morphologically and semantically can be summarized as in Figure 2.6.

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

Figure 2.6 Unity and PoS-system

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

[8] A further reflex of transnumerality, as noted in Rijkhoff (1993) is that verbal agreement with plural (or rather: collective) subjects is often singular. This is under certain conditions also true for Turkish (see Lewis 1967/1985: 246).

2.7 Pervasiveness

2.7.1 *Introduction*

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

2.7.2 *Levels of analysis*

2.7.2.1 *Introduction* The PoS-hierarchy and the resulting classification presented in Section 2.2 are based on a consideration of the behaviour of stems, both basic and derived. A number of recent papers, Haig (2006) being the first of these, have argued for considering the issue of flexibility at successive levels of morphosyntactic analysis. Haig (2006) more specifically argues for a principle of increasing categorization, i.e. decreasing flexibility, and Lehmann (2008) pursues this same issue independently. The levels of analysis that may be considered include at least the root, the basic stem, the derived stem, and the morphosyntactic word. The distinction between root and basic stem is especially relevant in those languages that manifest dependent roots, i.e. lexical roots that can only occur in combination with other roots or derivational affixes, and will not be considered here for lack of data from a substantial sample of languages.[9] The step from basic stems to derived stems is relevant to the relation between lexicon and syntax that is the central concern of this article, and will be addressed in Section 2.7.2.2. Once inserted into a morphosyntactic position, basic and derived stems acquire the status of morphosyntactic words that are part of phrases and clauses. I will limit the discussion of these to flexibility in the use of phrases as predicates in Section 2.7.2.3, and flexibility of the use of dependent clauses in various functional slots in Section 2.7.2.4.

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

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

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

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

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

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

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

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

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

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

Hengeveld (1992b) shows that languages with flexibility in their PoS-system, i.e. types 1 through 3/4, consistently allow possessive phrases to act as a non-verbal predicate, while all languages with some degree of rigidity consistently never allow

this. Differentiated languages sometimes do and sometimes do not allow the predicative use of possessive phrases.

A further issue that is of interest in relation to non-verbal predication is the way in which languages treat their non-verbal predicates from a formal perspective. Compare the following examples from Lango (Noonan 1992: 45) and !Xũ (Köhler 1981: 599):

(25) *Mán ʻgwôk*
DEM 3SG.dog.HAB
'This is a dog.'

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

In both examples a non-verbal predicate is used without the intervention of a copula, yet there is an important difference between the two. In Lango non-verbal predicates are inflected in precisely the same way as verbal predicates. In !Xũ, while verbal predicates are inflected regularly, the predicate in non-verbal predications is simply juxtaposed with its argument. In Hengeveld (1992b) the first strategy is called ∅1, the second ∅2.

One would expect the ∅1 strategy, which makes no difference between types of predicates, to be more typical of flexible PoS-systems, and this is indeed the case. Hengeveld (1992b) found that the ∅1 strategy is never used in languages with PoS-systems of types 5/6 through 7, whereas both ∅1 and ∅2 may be found in less rigid systems.

2.7.2.4 *Subordination* A remarkable fact about certain flexible languages is that the flexibility they exhibit in their PoS-system shows up in their subordination system as well. Consider the following examples from Turkish (Göksel and Kerslake 2005: 423–4), a type 2/3 language:

(27) *Orhan-ın bir şey yap-ma-y-acağ-ı belli-y-di-∅*
Orhan-GEN INDEF thing do-NEG-EV-NML.IRR-3SG.POSS obvious-EV-PST-3SG
'It was obvious that Orhan wouldn't do/wasn't going to do anything.'

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

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

[10] The differences in form are phonologically conditioned.

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

Van Lier (2009) (see also van Lier 2006; Hengeveld and van Lier 2008) shows that languages with PoS-systems 1 through 2/3 always have at least some subordinate construction that is flexible as well, though possibly with a lower degree of flexibility than the flexible PoS in those languages. This ties in neatly with the more general observation made in Section 2.7.2.1 that flexibility decreases as complexity increases.

2.7.3 *Further functions*

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

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

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

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

2.7.4 *Correlations*

The properties of flexible languages that follow from the fact that the flexibility/rigidity of basic PoS-systems may be extended to other areas of the grammar and the lexicon can be summarized as in Figure 2.7.

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

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

FIGURE 2.7 Pervasiveness and PoS-system

2.8 Conclusions

The cumulative results of our research can now be listed as in Figure 2.8. In this table, numbers following parameters refer to the relevant sections in this paper. Vertical bold lines show the relevant cut-off points between contiguous PoS-systems. These cut-off points just by themselves provide evidence for the relevance of the distinctions between all PoS-systems, including all intermediate ones, up to the distinction between type 5 and 5/6. Horizontal bold lines indicate which group of properties holds for a certain basic PoS system plus the next intermediate one. The groups of properties thus identified are cumulative, i.e. the properties under the highest horizontal bold line (i.e. all properties) hold for languages of types 1–1/2, the properties under the next horizontal bold line hold for languages of types 2–2/3, etc.

What can be learned from this inventory is that:

- the more flexible a language is in its use of lexemes, the more rigid it is in its syntax and morphology;
- the more flexible a language is in its use of lexemes, the more resistant it is to fusional morphology;
- the more flexible a language is in its use of lexemes, the more it lacks intrinsic lexical features, be they morphological or semantic in nature;
- the more flexible a language is in its use of lexemes, the more it is flexible in its use of phrases and clauses.

	1	2	3	4	5	6	7
Fusion—Head Predicate Phrase (2.5)	N	Y/N					
Conjugation classes (2.6.2)	N	Y/N					
Intrinsic transitivity (2.6.3)	N	Y/N					
Non-verb der. from flexible source (2.7.2.2)	Y	N					
Fusion—Head Referential Phrase (2.5)	N	Y/N					
Modifier der. from flexible source (2.7.2.2)	N	Y	N				
Predicate initial or final position (2.4.2)	Y	Y/N					
Overt marking of semantic functions (2.4.2)	Y	Y/N					
Declination classes (2.6.2)	N	Y/N					
Intrinsic number (2.6.3)	N	Y/N					
Flexible dependent clauses (2.7.2.4)	Y	Y/N					
Flex. loc. and temp. modifiers (2.7.3)	Y	Y/N					
Fusion—Modifier Referential Phrase (2.5)	N	Y/N					
Fixed order of head and modifier (2.4.3)	Y	Y/N					
Mann. adv. der. from flex. source (2.7.2.2)	N	Y	N				
Predicability of possessive phrase (2.7.2.3)	Y	N					
SVCs introducing participants (2.7.3)	N	Y/N					
Ø1-strategy (2.7.2.3)	Y/N	N					
	1	2	3	4	5	6	7

Figure 2.8 Grammatical and lexical properties of languages with different PoS-systems

A typical flexible language thus:

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

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

2.9 Appendix: samples used in the studies reported on

Source / *Language*	Fleur (1999)	Hengeveld (1992b)	Hengeveld (2007)	Hengeveld et al. (2004)	Hengeveld and Valstar (2010)	vanLier (2009)	Naeff (1998)	Rijkhoff (2004)	Rijkhoff (2003)	Smit (2007)
!Xũ		●								
Abkhaz	●	●	●	●	●	●	●	●	●	
Abun						●				
Alamblak			●	●	●	●		●	●	
Arabic, Egyptian		●								
Arapesh, Mountain			●	●	●	●		●	●	
Babungo		●	●	●	●	●	●	●	●	●
Bambara		●	●	●	●	●		●	●	
Basque	●	●	●	●	●	●	●	●	●	
Berbice Dutch	●		●	●	●	●	●	●	●	
Burmese								●	●	
Burushaski	●	●	●	●	●	●	●	●	●	
Cayuga								●	●	
Chinese, Mandarin		●	●	●	●	●	●	●	●	
Chukchi	●	●						●	●	
Dhaasanac						●				
Dutch		●						●	●	●
English										●
Galela								●	●	
Garo			●	●	●	●				

Georgian				●		●			●	
Gooniyandi						●				
Guaraní	●	●	●	●	●	●		●	●	
Gude	●		●	●	●			●	●	
Hausa		●								
Hdi						●				
Hittite			●	●	●			●	●	
Hixkaryana		●	●	●	●	●	●	●	●	
Hungarian	●	●	●	●	●	●		●	●	
Hurrian				●	●			●	●	
Ika								●	●	
Itelmen			●	●	●	●				
Jamaican Creole		●								
Japanese				●		●				
Kambera						●				
Kayardild	●		●	●	●	●	●	●	●	
Ket	●	●	●	●	●	●		●	●	●
Kharia						●				
Kisi			●	●	●	●		●	●	
Koasati			●	●	●	●		●	●	
Korean								●	●	
Krongo	●	●	●	●	●	●		●	●	
Lango	●	●	●	●	●	●	●	●	●	●
Lavukaleve						●				
Ma'di						●				
Mam		●								
Miao		●	●	●	●	●	●	●	●	
Mundari			●	●	●		●			●

Source / *Language*	Fleur (1999)	Hengeveld (1992b)	Hengeveld (2007)	Hengeveld et al. (2004)	Hengeveld and Valstar (2010)	vanLier (2009)	Naeff (1998)	Rijkhoff (2004)	Rijkhoff (2003)	Smit (2007)
Nahali		●								
Nama	●		●	●	●	●	●	●	●	
Nasioi		●	●	●	●			●	●	
Navaho	●	●	●	●	●		●			
Ngalakan		●	●	●	●			●	●	
Ngiti			●	●	●			●	●	●
Ngiyambaa		●								
Nivkh	●	●	●	●	●	●			●	
Nung	●		●	●	●	●		●	●	
Nunggubuyu			●	●	●	●		●	●	
Oromo			●	●	●		●	●	●	
Paiwan			●	●	●	●				
Pipil	●	●	●	●	●	●	●	●	●	
Polish	●		●	●	●	●				
Quechua, Imbabura	●	●	●	●	●	●	●	●	●	●
Sahu			●		●					
Samoan	●		●	●	●	●		●		●
Santali						●				
Sarcee								●	●	
Slave						●				
Sumerian		●	●	●	●			●	●	
Tagalog		●		●		●				

Tamil	●	●	●	●	●	●	●	●	●	
Thai		●				●				
Tidore				●						
Tsou								●	●	
Turkish	●	●	●	●	●	●	●		●	●
Tuscarora	●		●	●	●	●	●			
Vietnamese		●						●	●	
Wambon	●		●	●	●	●	●	●	●	
Warao			●	●	●	●				●
West Greenlandic	●	●	●	●	●	●		●	●	
Yagaria		●								
Yessan-Mayo		●								