

Parts-of-speech systems and lexical subclasses

Kees Hengeveld

Amsterdam Center for Language and Communication, University of Amsterdam

Marieke Valstar

Alkwin Kollege, Uithoorn

This paper shows that the potential presence of lexical subclasses in a language can be predicted from its parts-of-speech system. The classification of parts-of-speech systems presented in Hengeveld (1992a) is used as a starting point, and the notion of lexical subclass is restricted to those instances that are morphologically based. The investigation of a 50-language sample reveals that the functional flexibility of a lexical class that may occur in various syntactic slots prevents it from manifesting morphologically based subclasses. Lexical classes specialized for certain syntactic positions, on the other hand, do allow morphological subclassification.

1 Introduction¹

This paper investigates the implicational relations between the parts-of-speech (PoS) system of a language and the presence of lexical subclasses (declination classes, conjugation classes) in that language. The paper is organized as follows: §2 presents the language sample on which this study is based. In §3 we summarize the approach to parts-of-speech systems advocated in Hengeveld (1992a). This summary provides the necessary background for §4, which contains a presentation of our hypotheses concerning the relation between the parts-of-speech system of a language and the presence of lexical subclasses in the lexicon of that language. Our definition of a lexical subclass is given in §5, where we will restrict the use of this notion to morphological (as opposed to phonological and semantic) systems of lexical classification. We are then ready to present the data collected on the basis

¹ We are grateful to two anonymous reviewers of *Linguistics in Amsterdam* for their insightful comments. Abbreviations used: 1 = first person, 2 = second person, 3 = third person, ART = article, CONN = connective, COP = copula, DECL = declarative, F = feminine, GEN = genitive, INF = infinitive, IPFV = imperfective, IRRAT = irrational, LOC = locative, M = masculine, N = neuter, OBJ = object, POSS = possessive, PRES = present, PROGR = progressive, PST = past, SBJ = subject, SG = singular, TR = transitive.

of these definitions from the sample languages in §6. The results of the testing of the hypotheses are presented in §7, and our conclusions are given in §8.

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 satisfy 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 - it has been applied to Ruhlen's (1987) classification of the world's languages. We refer to Rijkhoff et al. (1993) for further details.

Within the restrictions of the genetic criterion the sample also represents maximal geographic diversity. Where possible, we have selected languages that are not spoken in contiguous areas.

Within the restrictions of the genetic and the geographic criterion the sample represents maximal typological diversity as well. Given our specific research question we have made sure that among the languages selected there are representatives of all major parts-of-speech systems as distinguished in section 3.

The criteria described above can only be met if adequate language descriptions are available. This is not the case. Data are insufficient or lacking for three out of the 50 languages that should be selected according to the genetic criterion (Etruscan, Meroitic, Nahali). The actual sample thus contains 47 languages.

The above description makes clear that the sample we use is not a random sample. Such a sample cannot be used in those cases in which one has to control the sample for a specific typological parameter, as is the case in this study. We are interested in the way in which languages exhibiting a range of parts-of-speech systems behave as regards lexical subclassification, and therefore had to use our pre-existing knowledge concerning the parts-of-speech systems of the languages concerned in setting up the sample.

An objection that might be raised against the sample is that the number of languages for which the hypotheses are relevant is rather small. This is a result of the fact that this study is part of a larger project in which a whole range of morphosyntactic properties of languages are correlated with their parts-of-speech systems. The results of that project are summarized in Hengeveld (forthc.). Given the overall aims of that project the sample has to be kept stable over all of its subparts. When presenting our results in section 7 we will clearly indicate the number of sample languages on which our conclusions are based.

Table 1: The sample

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

3 Parts-of-speech²

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

	head	modifier
predicate phrase	1	4
referential phrase	2	3

Figure 1: Lexemes and functions

Figure 1 shows that the functional positions 1-4 are based on two parameters, one involving the opposition between predication and reference, the other between heads and modifiers. Together, these two parameters define the following four functions: head of a predicate phrase (1), head of a referential phrase (2), modifier in a referential phrase (3), and modifier in a predicate phrase (4). 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 1: verbs like *sing* are used as heads of predicate phrases; nouns like *girl* as heads of referential phrases; adjectives like *tall* as modifiers in referential phrases; and manner adverbs like *beautifully* as modifiers in predicate phrases. Crucially, none of the content lexemes in (1) could be used directly in another function, i.e. without morpho-syntactic adaptation. Thus, in this example there is a one-to-one relation between function and lexeme class. Parts-of-speech systems of this type are called differentiated, and the lexical classes can all be said to be specialized for a certain propositional function.

There are other parts-of-speech systems in which there is no one-to-one relation between the four propositional functions identified and the lexeme

² This section is largely based on earlier summaries of the model, such as the ones in Hengeveld (2007, forthc.) and Hengeveld & van Lier (2008, 2009).

classes available. These systems are of two types. In the first type, a single class of lexemes is used in more than one propositional function. Such lexeme classes, and the parts-of-speech systems in which they appear, are called *flexible*. The second type is called *rigid*. Rigid systems resemble differentiated systems to the extent that both consist only of lexemes classes that are specialized, i.e. dedicated to the expression of a single function. However, rigid systems are characterized by the fact that they do not have four lexeme classes; one for each of the four propositional functions. Rather, for one or more functions a lexeme class is lacking. The following examples illustrate the difference between these flexible and rigid parts-of-speech systems. In Turkish (Göksel & Kerslake 2005: 49) the same lexical item may be used indiscriminately as the head of a referential phrase (2), as a modifier within a referential phrase (3), and as a modifier within a predicate phrase (4):

- (2) güzel-im
beauty-1.POSS
'my beauty'
- (3) güzel bir köpek
beauty ART dog
'a beautiful dog'
- (4) Güzel konuş-tu-Ø.
beauty speak-PST-3.SG
'S/he spoke well.'

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

- (5) Álímì bìitì.
be.cold.M.IPFV water.
'The water is cold.'
- (6) bìitì ñ-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 *ñ-* and its allomorphs. This is the general relativizing strategy in Krongo, as illustrated by the following examples (Reh 1985: 256):

As Figure 2 shows, Turkish and Krongo are similar in that they have two main classes of lexemes. They are radically different, however, in the extent to which one of these classes may be used in the construction of predication: 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. Note that for a lexeme class to classify as flexible, the flexibility should not be a property of a subset of items, but a general feature of the entire class.

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

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

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

The hierarchy in (10), combined with the distinction between flexible, differentiated, and rigid languages, predicts a set of seven possible parts-of-speech systems, which is represented in Figure 3. As this figure shows, it is predicted that languages can display three different degrees of flexibility (systems 1-3), three different degrees of rigidity (systems 5-7), or can be differentiated (type 4). Of the languages discussed earlier Turkish would be a type 2 language, English a type 4 language, and Krongo a type 6 language. Note that we 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.

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

Figure 3: Parts-of-speech systems

In addition to the seven types listed in Figure 3, there are so-called intermediate systems, showing characteristics of two systems that are contiguous in Figure 3.

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

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

In flexible languages the most common source for such an intermediate status is that derived stems show a lower degree of flexibility than basic stems. In rigid languages the most common source of an intermediate status is the existence of small, closed classes of lexemes at the fringe of the system. Figure 4 (see also Smit 2007) shows the full set of possible systems, including the intermediate ones.

For further details on and argumentation for the approach to parts-of-speech systems outlined in this section see Hengeveld, Rijkhoff & Siewierska (2004).³

4 Hypotheses

On the basis of the preceding classification of parts-of-speech systems, we may now formulate the following hypotheses concerning the occurrence of lexical subclasses in a language in relation to its parts-of-speech system. Consider first the general hypothesis given in (11):

(11) General hypothesis

The higher the degree of morphological unity (i.e. the absence of intrinsic subclasses triggering specific morphological processes) of a lexical class is, the higher its degree of applicability in various syntactic slots is. Intrinsic lexical subclasses are therefore not expected to occur in flexible languages.

The general idea that forms the basis for this hypothesis is that lexical subclassification is a feature of lexemes that are used in specific syntactic slots. Thus, declination classes are a feature of lexemes when used as the head of a referential phrase, and conjugation classes are a feature of lexemes when used as the head of a predicate phrase. In flexible languages, intrinsic lexical subclasses based on the use of lexemes in specific syntactic slots would combine functional flexibility with formal inflexibility. Although not logically impossible, this combination is highly unlikely, since it places a heavy burden on language production, as it would require the speaker to use different subclassifications for the same lexeme depending on the function in which it is used.

³ Hengeveld & van Lier (2008, 2009) propose a 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 that have been attested. Since these adaptations do not affect the points and generalizations made in this paper, we refrain from presenting this new approach here.

We will investigate this hypothesis for two syntactic slots: the head slot of referential phrases and the head slot of predicate phrases. With respect to these slots, we derive the two specific hypotheses in (12) from the general hypothesis in (11):

(12) Specific hypotheses

- a. In languages with a true class of verbs (types 2-7), the lexical elements that are used as the head of a predicate phrase may display intrinsic class distinctions (henceforth ‘conjugation classes’), whereas in languages with a flexible class of lexical elements that may be used as the head of a predicate phrase (types 1-1/2), such distinctions are not expected to occur.
- b. In languages with a true class of nouns (types 3-6/7), the lexical elements that are used as the head of a referential phrase may display intrinsic class distinctions (henceforth ‘declination classes’), whereas in languages with a flexible class of lexical elements that may be used as the head of a referential phrase (types 1-2/3), such distinctions are not expected to occur.

Note that we use the terms ‘declination class’ and ‘conjugation class’ in a broad sense for all intrinsic class distinctions made within classes of lexical elements that are used as the heads of predicate phrases and referential phrases respectively.

Before confronting the hypotheses in (12) with the data from the sample languages, we will briefly explain how we interpret the notion of intrinsic lexical subclass.

5 Lexical subclasses

5.1 Definition

By lexical subclasses we understand intrinsic subclasses of a lexeme type that trigger specific morphological processes. This means we are interested in morphological systems (Corbett 1991: 34) of lexical subclassification only. In this type of system, membership in a subclass is an intrinsic property of the lexemes involved, which is crucial to our hypotheses. In morphological systems, the membership of a lexeme of a specific lexical class has to be stored. By using these definitions we exclude semantic and phonological systems of lexical subclassification. In these types of system, membership in a subclass is fully predictable on the basis of the shape or the meaning of the lexemes involved,

and it is not an intrinsic property of the lexemes themselves. In order to clarify these distinctions, we will briefly illustrate the three types of system in the next section.

5.2 *Semantic, phonological and morphological systems*

5.2.1 *Semantic systems*

A semantic system is a system in which membership in a subclass is determined entirely by the meaning of a lexeme. A language in which a semantic system can be found is Abkhaz, for example. In Abkhaz, a type 4 language, a semantic system is present in verbs. At first sight, there seem to be two verb classes in Abkhaz: verbs which take the declarative marker -yt' and verbs which take the declarative marker -p'. Dynamic verbs take the former, and stative verbs take the latter suffix. The two construction types are illustrated in (13)-(14) (Spruit 1986: 99, 98):

- (13) Də-s-táa-wa-yt'.
3.SG.M.SBJ-1.SG.OBJ-visit-PROGR-DECL
'He visits me.'
- (14) Yə-s-taxə-w-p'.
3.SG.IRRAT-1.SG-want-PRES-DECL
'I want it.'

Upon closer inspection, however, it turns out that the choice for a particular marker is dictated by the meaning a speaker wants to transmit. So in some cases both declarative endings are possible, depending on the intended meaning. This is illustrated in (15)-(16) (Spruit 1986: 95, 96). In (15) the action of sitting down is expressed, whereas in (16) the state of sitting is expressed:

- (15) D-t'^wa-wá-yt'.
3.SG.M-sit-PROGR-DECL
'He sits down.'
- (16) D-t'^wá-w-p'.
3.SG.M-sit-PRES-DECL
'He is sitting.'

These examples show that the classification of a verb in Abkhaz can be predicted from its meaning: it is a semantic rather than an intrinsic property of the verb.

Another semantic system can be found in Ket, a type 3 language, with nouns when used as the head of referential phrases. Nouns in Ket take either masculine, feminine or non-human (animate or inanimate) markers. The marker selected for a noun corresponds directly with its meaning, in the sense that all masculine, feminine and non-human nouns are nouns denoting men, women and non-human entities respectively. But again, the choice for a particular marker is dictated by the meaning a speaker wants to transmit, as a result of which more than one marker is possible in some cases, as illustrated in (17) and (18) (Werner 1997: 88):

- (17) 1o·ks' 'tree' (animate), 'stick' (inanimate)
 (18) ⁴qal 'granddaughter' (feminine), 'grandson' (masculine)

Thus, Ket displays a semantic system rather than a morphological system: class membership of a noun can be predicted on the basis of its meaning and is not an intrinsic property of the lexeme itself.

5.2.2 Phonological Systems

A phonological system is a system in which membership in a subclass is determined entirely by the form of a lexeme. A phonological system can be found in Turkish, a type 2/3 language. Suffixes added to verbs in Turkish take various shapes, depending on the quality of the last vowel of the verb stem. This is illustrated for the progressive suffix in (19), which has four allomorphs. The same type of vowel harmony may be observed with lexemes occurring as the head of referential phrases, as illustrated for the genitive suffix in (20) (Lewis 1967: 109, 30-31):

- | | | | | |
|------|------------|-------------|-------------|-------------|
| (19) | agel-iyor | b al-ıyor | c gör-üyor | d koş-uyor |
| | come-PROGR | take-PROGR | see-PROGR | run-PROGR |
| (20) | agece-nin | b tarla-nın | c ölçü-nün | d korku-nun |
| | night-GEN | field-GEN | measure-GEN | fear-GEN |

Thus, Turkish displays a phonological system rather than a morphological system: class membership of a lexeme can be predicted on the basis of its form by general phonological rules, and is not an intrinsic property of the lexeme itself.

5.2.3 Morphological systems

The systems we are looking for are systems not based on semantic and phonological distinctions: morphological systems (Corbett 1991: 34) in which class membership is an intrinsic property of the lexeme itself, i.e. not predictable on the basis of meaning or form. A morphological system can be found in Kisi, a type 5 language, with lexemes occurring as the head of a predicate phrase. Verbs in Kisi can be divided into two classes: the so-called ‘regular’ and ‘irregular’ verbs. This latter class has a number of subclasses, of which we only illustrate one here. The stem-vowel of regular verbs is the same in all tense/aspect verb forms, while the stem-vowel of irregular verbs changes in certain tense/aspect verb forms (Childs 1995: 220-1)

- (21) a Regular (Affirmative)
- | | | |
|-------------|--------------------|----------------------|
| Stem | <i>cimbu</i> | ‘leave’ |
| Habitual | <i>ò cimbù</i> | ‘she usually leaves’ |
| Perfective | <i>ò cimbú</i> | ‘she left’ |
| Perfect | <i>ò cimbú nîŋ</i> | ‘she has now left’ |
| Hortative | <i>ò cimbú</i> | ‘she ought to leave’ |
| Past Habit | <i>óó cimbù</i> | ‘she used to leave’ |
| Subord | <i>mbò cimbù</i> | ‘that she leaves’ |
| Past Subord | <i>mbó cimbù</i> | ‘that she left’ |
- b Irregular (Affirmative – o/e-group)
- | | | |
|-------------|------------------|---------------------|
| Stem | <i>kiol</i> | ‘bite’ |
| Habitual | <i>ò kìàl</i> | ‘she usually bites’ |
| Perfective | <i>ò kèl</i> | ‘she bit’ |
| Perfect | <i>ò kèl nîŋ</i> | ‘she has bitten’ |
| Past Habit | <i>óó kìàl</i> | ‘she used to bite’ |
| Hortative | <i>ò kìól</i> | ‘she ought to bite’ |
| Subord | <i>mbò kìàl</i> | ‘that she bites’ |
| Past Subord | <i>mbó kìàl</i> | ‘that she bit’ |

In order to use the appropriate forms, these have to be stored, i.e. class membership is a property of the lexeme itself.

Another morphological system can be found in Polish, a type 4 language, with lexemes occurring as the head of referential phrases. Polish has masculine, feminine and neuter nouns. The gender of a noun is not predictable on the basis of meaning or form. The following examples illustrate this (Teslar 1953: 255, 266, 270):

- (22) a. Masc *kraj* ‘country’
- | | |
|--------------|-----------------|
| Nominative | <i>kraj</i> |
| Genitive | <i>kraj-u</i> |
| Dative | <i>kraj-owi</i> |
| Accusative | <i>kraj</i> |
| Vocative | <i>kraj-u</i> |
| Instrumental | <i>kraj-em</i> |
| Locative | <i>o kraj-u</i> |
- b. Neuter *okno* ‘window’
- | | |
|--------------|----------------|
| Nominative | <i>okn-o</i> |
| Genitive | <i>okn-a</i> |
| Dative | <i>okn-u</i> |
| Accusative | <i>okn-o</i> |
| Vocative | <i>okn-o</i> |
| Instrumental | <i>okn-em</i> |
| Locative | <i>o okn-e</i> |
- c. Feminine *ziemia* ‘earth, ground’
- | | |
|--------------|-----------------|
| Nominative | <i>ziem-ia</i> |
| Genitive | <i>ziem-i</i> |
| Dative | <i>ziem-i</i> |
| Accusative | <i>ziem-ię</i> |
| Vocative | <i>ziem-io</i> |
| Instrumental | <i>ziem-iq</i> |
| Locative | <i>o ziem-i</i> |

Polish exhibits a morphological system, since for each noun one has to know which subclass it belongs to.

6 The data

We are now ready to establish for each of the languages in our sample whether it exhibits lexical subclasses of the morphological type for lexemes that are used as heads of referential phrases (declination classes) and predicate phrases (conjugation classes) respectively. In Table 3 we present our findings. ‘Y’ means ‘morphological system attested’, and ‘N’ means ‘no morphological system attested’. Systems which are partly morphological and partly semantic and/or phonological are counted as morphological systems.

Table 3: The data

Language	Parts-of-speech system	Conjugation classes	Declination classes
Samoa	1	N	N
Guaraní	1/2	N	N
Mundari	1/2	N	N
Hurrian	2	N	N
Quechua	2	N	N
Warao	2	N	N
Turkish	2/3	Y	N
Ket	3	Y	N
Miao	3	N	N
Ngiti	3	Y	N
Burushaski	3/4	N	N
Lango	3/4	N	Y
Abkhaz	4	N	N
Arapesh	4	Y	Y
Babungo	4	N	Y
Bambara	4	Y	N
Basque	4	N	N
Itelmen	4	Y	N
Hittite	4	Y	Y
Hungarian	4	Y	N
Nama	4	? ⁴	Y
Ngalakan	4	Y	N
Polish	4	Y	Y
Nasioi	4/5	N	N
Oromo	4/5	N	Y
Pipil	4/5	N	N
Sahu	4/5	N	N
Sumerian	4/5	Y	N
Alamblak	5	Y	Y
Berbice Dutch	5	N	N
Kayardild	5	Y	Y
Kisi	5	Y	N
Koasati	5	Y	N
Paiwan	5	Y	N
Wambon	5	N	N
Chinese	5/6	N	Y
Garó	5/6	N	N
Gude	5/6	N	Y
Nung	5/6	N	N
Tamil	5/6	Y	N
West Greenlandic	5/6	N	Y
Gilyak	6	N	N
Hixkaryana	6	Y	N
Krongo	6	N	Y
Navaho	6	? ⁵	N
Nunggubuyu	6	? ⁶	N
Tuscarora	6/7	N	N

⁴ It is not clear whether the nature of the difference between so called active and neuter verbs in Nama is morphological or semantic.

⁵ The system of verbal subclassification in Navaho is based upon the classifiers that the verbs take. Because the exact function of these classifiers is not clear, it is not possible to decide whether the system is a morphological or a semantic one.

⁶ There is a system of verbal subclassification in Nunggubuyu, but it is impossible to say where the phonological/semantic systems end and the morphological system starts.

7 Results

7.1 Testing the hypotheses

The data in Table 3 fully confirm the specific hypotheses given in (12). The first hypothesis predicts that intrinsic conjugation classes are not found in languages with PoS-systems 1-1/2. The languages in the sample behave in an even stricter way than we predicted, since intrinsic conjugation classes are not found in languages with PoS-systems up to and including 2. These results are summarized in Figure 5. The figures in each cell indicate the number of languages exhibiting the given combination of features.

	intrinsic conjugation classes	no intrinsic conjugation classes
PoS-system 1-2	- (0)	+ (6, e.g. Warao)
PoS-system 2/3-7	+ (18, e.g. Hittite)	+ (20, e.g. Basque)

Figure 5: Conjugation classes

The second hypothesis predicts that intrinsic declination classes are not found in languages with PoS-systems 1-2/3. Again, the languages in the sample behave in an even stricter way than predicted, since intrinsic declination classes are not found in languages with PoS-systems up to and including 3. These results are summarized in Figure 6.

	declination classes	no declination classes
PoS-system 1-3	- (0)	+ (10, e.g. Hurrian)
PoS-system 3/4-6/7	+ (13, e.g. Alamblak)	+ (24, e.g. Garo)

Figure 6: Declination classes

An interesting question is why the languages in the sample would in both cases confirm to the hypothesis for one additional PoS system, PoS system 2 for the first hypothesis and PoS system 3 for the second hypothesis. The most likely explanation for this is that (i) languages of type 2 generously allow the use of non-verbs in predicative position, often without the intervention of a copula, thus creating functional overlap between verbs and non-verbs as heads of predicative phrases; (ii) languages of type 3 generously allow the use of modifiers in contextually licensed headless referential phrases, thus creating functional overlap between nouns and modifiers as the central elements of referential phrases.

7.2 Other observations

Apart from the results we predicted in our hypotheses, Table 3 shows a correlation that we did not predict: only in languages with PoS-systems 4 up to and including 5 do we find the presence of both intrinsic conjugation classes and intrinsic declination classes. The languages involved are Arapesh, Hittite and Polish (type 4); and Alamlak and Kayardild (type 5). This finding is summarized in Figure 7.

	both	one or none
PoS-system 1-3/4, 5/6-7	— (not attested)	+ (e.g. Ngiti)
PoS-system 4-5	+ (e.g. Polish)	+ (e.g. Pipil)

Figure 7: Intrinsic conjugation and declination classes

Although we do not have an explanation for this result, it does indicate that lexical subclassification is typical of languages with a differentiated parts-of-speech system.

8 Conclusions

The investigation of a 50-language sample confirms our hypotheses concerning the occurrence of intrinsic conjugation and declination classes in relation to the parts-of-speech system of a language. The flexibility of a parts-of-speech system imposes restrictions on the degree of specialization for syntactic slots that

lexical items may exhibit. Our additional finding that the combination of intrinsic conjugations and intrinsic genders is only found in languages with parts-of-speech systems with a high degree of differentiation provides further confirmation for the relation between specialization and subclassification.

References

a. Descriptions of sample languages

Abkhaz

Hewitt, B.G. (1979). *Abkhaz* (Lingua Descriptive Studies 2). Amsterdam: North Holland.

Spruit, A. (1986). *Abkhaz studies*. Dissertation, University of Leiden.

Alamblak

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

Arapesh, Mountain

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

Babungo

Schaub, W. (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.

Chinese, Mandarin

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

Garó

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

Guaraní

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

Gude

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

Hittite

Friedrich, J. (1974). *Hethitisches Elementarbuch. I: Kurzgefasste Grammatik*. 3. Auflage. 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

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

Hurrian

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

Itelmen

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

Kayardild

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

Ket

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

Kisi

Childs, G. Tucker (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: University of Nebraska Press.

Krongo

Reh, Mechthild (1985). *Die Krongo-Sprache (nìino 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.

Miao

Harriehausen, Bettina (1990). *Hmong Njua* (Linguistische Arbeiten 245). Tübingen: Max Niemeyer.

Mundari

Cook, W.A. (1965). *A descriptive analysis of Mundari*. Dissertation, Georgetown University.

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

Nama

Hagman, R.S. (1974). *Nama Hottentot Grammar*. Dissertation, Columbia University.

Rust, F. (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. & William Morgan, Sr. (1987). *The Navajo language: a grammar and colloquial dictionary*. Revised edition. Albuquerque: University of New Mexico Press.

Ngalakan

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

Ngiti

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

Nivkh

Gruzdeva, Ekaterina (1998). *Nivkh* (Languages of the World: Materials 111). München: Lincom.

Mattissen, Johanna & Werner Drossard (1998). *Lexical and syntactic categories in Nivkh* (Theorie des Lexicons: Arbeiten des Sonderforschungsbereich 282, 85). Düsseldorf: Heinrich Heine Universität.

Nung

Saul, Janice E. & Nancy Freiberger Wilson (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

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

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 edition. Edinburgh: Oliver and Boyd.

Quechua

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

Sahu

Visser, L.E. & C.L. Voorhoeve (1987). *Sahu-Indonesian-English dictionary and Sahu Grammar Sketch* (Verhandelingen van het Koninklijk Instituut voor Taal-, Land- en Volkenkunde 126). Dordrecht: Foris.

Samoan

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

Sumerian

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

Tamil

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

Turkish

Göksel, Asli & Celia Kerslake (2005). *Turkish. A comprehensive grammar*. London/New York: Routledge.

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

Lewis, G.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

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

Warao

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

West Greenlandic

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

b. Other references

- Corbett, Greville (1991). *Gender* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.
- Hengeveld, Kees (1992a). Parts of speech. In: Michael Fortescue, Peter Harder & Lars Kristoffersen (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.
- Hengeveld, Kees, Jan Rijkhoff & Anna Siewierska (2004). Parts-of-speech systems and word order. *Journal of Linguistics*, 40: 527-570.
- Hengeveld, Kees (2007). Parts-of-speech systems and morphological types. *ACLIC Working Papers*, 2.1: 31-48.
- Hengeveld, Kees (forthc.). Parts-of-speech system as a basic typological determinant. To appear in: Jan Rijkhoff & Eva van Lier (eds). *Flexible word classes: a typological study of underspecified parts-of-speech*. Oxford: Oxford University Press.
- Hengeveld & van Lier (2008). Parts of speech and dependent clauses in Functional Discourse Grammar. In: Ansaldo, Umberto, Jan Don & Roland Pfau (eds). *Parts of Speech: Descriptive tools, theoretical constructs*. *Studies in Language*, 32.3: 753–785.
- Hengeveld & van Lier (forthc.). The implicational map of parts-of-speech. To appear in: Andrej Malchukov, Michael Cysouw & Martin Haspelmath (eds). *Semantic Maps: Methods and Applications*. *Linguistic Discovery* 7.2.
- Rijkhoff, Jan N.M., Dik Bakker, Kees Hengeveld & Peter Kahrel (1993). A method of language sampling. *Studies in Language*, 17.1: 169-203.
- Ruhlen, Merritt (1987). *A guide to the world's languages, Vol.1: Classification*. Stanford: Stanford University Press.
- Smit, Niels (2007). Dynamic perspectives on lexical categorization. *ACLIC Working Papers*, 2.1: 49-75.