



UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Parts of speech and dependent clauses: A typological study

van Lier, E.H.

Publication date
2009

[Link to publication](#)

Citation for published version (APA):

van Lier, E. H. (2009). *Parts of speech and dependent clauses: A typological study*. [Thesis, fully internal, Universiteit van Amsterdam]. LOT.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

PARTS OF SPEECH

2

2.1. Introduction

Parts of Speech (henceforth PoS) is the traditional term used to refer to the major classes of words or lexemes that are distinguished in a language (Schachter & Shopen 2007: 1). The issue of lexical classification or categorization has a long and troublesome history in linguistics. Generally, the nature of categorization in human language has been and continues to be the subject of a debate that goes back to Aristototele and has raised the interest of philosophers, psychologists, and linguists alike. In the present chapter, I will not attempt to give an overview of the various conceptions of categorization that have been on and off stage at various periods in the history of these sciences¹. Rather, I will limit myself to a brief discussion of the problems that have been at stake in defining lexical categories within the sub-discipline of linguistics called *functional typology*². This is the topic of section 2.2.

Subsequently, in section 2.3, I go on to discuss in more detail one specific functional-typological approach to lexical categorization, namely the theory proposed by Hengeveld (1992 a, b; Hengeveld et al. 2004; Hengeveld & Rijkhoff 2005; Hengeveld & Van Lier 2008, 2009). I will explain that

¹ For a recent overview of this type, see Aarts 2006.

² For other overviews of approaches to lexical categorization, including approaches that do not belong to the functional-typological framework, the reader is referred to Bisang (fc.), Evans (2000), Haspelmath (2001), and Rijkhoff (2007).

the Hengeveldian approach provides a suitable basis for the present study, because it defines PoS classes in terms of the *propositional function(s)* they can fulfil in an utterance³. These PoS definitions will serve as a basis for the typology of PoS systems presented in Chapter 5.

However, Hengeveld's approach is certainly not without problems. In section 2.4 I discuss a number of objections that have been raised against this theory, most elaborately by Croft (2000, 2001, 2005). In general terms, these problems involve the tension between the language specific nature of lexical categories on the one hand, and their cross-linguistic comparability on the other hand. As such, they touch upon a fundamental discussion in linguistic typology concerning the (non-)existence of universal linguistic categories (see Haspelmath 2007, Newmeyer 2007).

Section 2.5 is specifically devoted to lexical flexibility, which for the time being I will define as the possibility to use the members of one or more lexeme class(es) of a language in more than one propositional function without needing any morpho-syntactic adaptation. First, I will review the discussion about lexical flexibility in recent literature. Subsequently, I present an approach to lexical flexibility that attempts to integrate insights of various earlier studies (see also Don & Van Lier, forthcoming). Finally, section 2.6 provides a summary of this chapter.

2.2. Functional-typological attempts at defining parts of speech

As the binary term implies, a *functional-typological* approach to the study of linguistics involves two basic ingredients: typology and functionalism. First, typology deals with cross-linguistic comparison, under the assumption that the structural variation displayed by languages across the word is not random, but is rather constrained by a set of principles that are called typological *universals*. Second, functionalism refers to the assumption that these universals of language structure can be explained in terms of language use, i.e. in terms of the semantic and pragmatic meanings expressed by these structures in human communication⁴.

Within the functional-typological literature, various approaches to lexical categorization have been proposed. These can be characterised in terms of the type of criteria on which the definitions of PoS classes are primarily based.

³ The term *propositional function* has been introduced by Croft (2001). Its use in the present study will be clarified in the course of this chapter.

⁴ For a more detailed characterization of the functional-typological approach the reader is referred to Croft (2003), or Cristofaro (2003: 5-9).

According to Givón (2001: 49) there are three different types of such criteria:

- (i) **Semantic criteria:** the kind of *meanings* that tend to be coded by words of a particular class.
- (ii) **Morphological criteria:** the kind of *bound morphemes* that tend to be affixed to words of a particular class.
- (iii) **Syntactic criteria:** the typical *syntactic slot* that words of a particular class tend to occupy.

To these, a fourth type may be added, namely:

- (iv) **Pragmatic criteria:** the kind of *discourse function* that tends to be coded by words of a particular class. (cf. Sasse 1993b:196)

Although most approaches acknowledge the relevance of all four types of criteria for the definition of lexical categories, there are clear differences in emphasis. First, there is the so-called *notional approach*, which takes the semantic meaning of lexemes as the most important property of PoS classes, i.e. the types of entities that groups of lexemes denote. Thus, nouns are defined as lexemes denoting *first-order entities* or objects and individuals. In contrast, verbs denote *second-order entities*, i.e. “*events, processes, states-of-affairs, etcetera, that are located in time and which are said to ‘occur’ or ‘take place’ rather than to ‘exist’*” (Lyons 1977: 443). Langacker (1987) provides a notional definition of nouns and verbs from the perspective of Cognitive Grammar. According to his definitions, nouns designate bounded *regions*, typically in the domain of space, and are therefore cognitively perceived as static and holistic. Verbs, on the other hand, designate *processes*, and are cognitively perceived as sequences of states evolving through time.

The obvious problem of the notional approach, as has been pointed out by many scholars (Dik 1989; Lehmann 1990, among others), is that the same type of entity may be expressed by lexemes that belong to different lexical classes, both within and across languages. The English word *destruction*, for instance, is a noun, even though it denotes an event rather than an object. Furthermore, property concepts such as ‘small’ are in some languages lexicalized in the form of a separate class of adjectives, while in other languages these concepts are expressed by verbs or nouns.

A second possibility is the *discourse approach*, advanced by Hopper and Thompson (1984, 1985). It defines the differences between PoS classes

in terms of their prototypical discourse function. Thus, a prototypical verb expresses event-occurrence, whereas a prototypical noun introduces a participant into the discourse. Whenever a lexeme does not express its prototypical discourse function, it will show reduced categoriality, meaning that it will have fewer possibilities to take morphological marking. However, as in the case of the notional approach, the discourse approach makes no principled distinction between universal discourse functions and the language-specific lexical categories that express them.

In an attempt to overcome the problem of the notional and discourse-based approaches, focus shifted more towards the *morphological* characteristics of PoS. For the classical languages, it had been noted long ago that lexical classes could be identified by means of the particular morphological categories for which they are specified. Typical nominal categories would include grammatical gender, number, definiteness and case distinctions. Verbs rather express categories like tense, mood, aspect, and person agreement. And adjectives take morphological markers expressing degrees of comparison and may display gender, number and/or case agreement with the head they modify⁵.

As Croft (1991: 86) observes, the morphological categories mentioned above are indeed “*tailored to*” a specific lexical class, because they express bits of meaning that are relevant to the semantic meaning or the pragmatic-syntactic function prototypically expressed by members of that class. Thus, aspectual distinctions are relevant to verbs, which prototypically express actions; number is relevant to nouns, which prototypically refer to countable objects; and comparative marking is relevant to adjectives, which typically express gradable property concepts. Similarly, agreement has the function of showing the relationships between typically verbal predicates and their typically nominal argument(s), or between typically nominal heads and their typically adjectival modifiers.

However, morphological criteria are problematic too, because they are no more than “*diagnostic*” (Beck 2002:14) or “*symptomatic*” (Langacker 2002: 60) for lexical categories. For one thing, not all PoS classes in all languages express the same type and amount of morphological categories. Therefore, morphological criteria do not provide a suitable basis for cross-linguistic

⁵ As far as (manner) adverbs are concerned, it seems that there are no typical inflectional classes associated with this lexeme class. This is probably related to the fact that the status of manner adverbs as a major word class is somewhat problematic in the first place (see Haspelmath 2001: 16544).

generalizations. In addition, morphological categories are problematic as *intra*-linguistic criteria as well, because words belonging to the same lexical class often do not display the full range of morphological categories belonging to that class.

Schachter (1985) tries to come up with more integrated definitions of PoS classes. He argues that, apart from morphological criteria, the possible *syntactic* function(s) of lexemes should be taken into account as well⁶. However, this combination of formal criteria still yields definitions of PoS classes that are language-specific rather than universal. At the same time, Schachter uses notional labels to refer to lexeme classes, which according to him, “*appropriately reflect universal semantic considerations*” (Schachter 1985: 4). This suggests that some kind of common prototypical meaning of the members of a lexeme class should be acknowledged and play a part in a universal theory of lexical categorization.

Anward et al. (1997: 171) also argue in favour of an approach to PoS definition that integrates semantic, morphological, and syntactic, as well as phonological criteria. However, they clearly state that “*given that such properties of constituents differ from one language to another at least in part, parts-of-speech classes will need to be set up separately for each language [...]*”. Again, this approach has obvious repercussions for the cross-linguistic comparability of PoS classes.

To sum up, there are several possible approaches to the definition of PoS classes, each with its own problems. These problems can be broadly divided into two types: either universal tendencies are defined without accounting for language-specific categories, or the other way around. In the next sections, I will discuss in more detail two recent functional-typological theories of PoS classification that deal with these problems in very different ways: Hengeveld’s syntactic-pragmatic approach (section 2.3) and Croft’s Universal-Typological approach (section 2.4). These sections, together with the present one, should make clear that there is no single answer to the question of whether and how the four types of criteria – pragmatic, semantic, syntactic, and morphological – should be combined into definitions of lexical categories within and across languages. As Bisang (fc.) puts it: “*There is no final solution appearing on the horizon*”.

⁶ See also Schachter & Shopen (2007), who explicitly distinguish between notional and grammatical criteria for defining PoS.

2.3 Hengeveld's syntactic-pragmatic approach

2.3.1 Slots and fillers

Hengeveld (1992) developed his PoS theory as part of his typological study of non-verbal predication. He defines PoS classes in terms of the syntactic slot or slots that their members can occupy in an utterance, without any further morpho-syntactic measures⁷. Hengeveld rules out semantics as a component of PoS definitions, because, as explained in the previous section, denotational meaning is not a water-tight criterion for lexical category membership. The expression of morphological categories is used as an analytic device only, which can help to identify possible candidates for lexeme classes in specific languages. It does not, however, play a role in the actual definition of PoS classes.

Working in the framework of *Functional Grammar* (henceforth FG; Dik 1989, 1997), Hengeveld distinguishes four possible syntactic slots in which lexemes may be inserted. Two of these slots are obligatorily filled in their respective phrases, namely:

- (i) The head of a predicate phrase and
- (ii) The head of a referential phrase

The other two slots are optionally filled in each of the two phrase types. They are:

- (iii) A modifier within a predicate phrase and
- (iv) A modifier within a referential phrase.

The four slots can thus be tabulated as in Figure 2.1:

	Head	Modifier
Predicate phrase	(i)	(iv)
Referential phrase	(ii)	(iii)

Figure 2.1: Syntactic slots as distinguished in Hengeveld (1992)

On the basis of their distribution over these four syntactic slots, the

⁷ As I will explain shortly, in more recent work (Hengeveld & Van Lier 2008, 2009) the syntactic slots are re-interpreted as propositional functions.

definitions of the major PoS classes verb, noun, adjective, and manner adverb are formulated as in (i):

- (i)
- (i) A *verb* is a lexeme that can be used, without further morpho-syntactic measures, as the head of a predicate phrase only (slot (i));
 - (ii) A *noun* is a lexeme that can be used, without further morpho-syntactic measures, as the head of a referential phrase (slot (ii));
 - (iii) An *adjective* is a lexeme that can be used, without further morpho-syntactic measures, as a modifier in a referential phrase (slot (iii));
 - (iv) A *manner adverb* is a lexeme that can be used, without further morpho-syntactic measures, as a modifier in a predicate phrase (slot (iv)).
- (Hengeveld 1992: 58; Hengeveld et al. 2004: 530).

In Figure 2.2 the four syntactic slots are represented once again; this time with the four lexeme classes appearing in the relevant slots (from Hengeveld et al. 2004: 530):

	Head	Modifier
Predicate phrase	verb	manner adverb
Referential phrase	noun	adjective

Figure 2.2: Syntactic slots and lexical categories

The definitions of the PoS classes in (i) give rise to a number of issues that require some clarification. First, there is quite a bit more to say about the proviso ‘*without further morpho-syntactic measures*’. For a proper understanding of what constitute such a measure, I find it useful to refer to a distinction made by Croft (1991, 2001), between *structural coding and behavioural potential*⁸. The first term, *structural coding*, refers to those morpho-syntactic markers that specifically indicate the syntactic or propositional function of a linguistic unit (a lexeme, in the case at hand). The other type of marking, *behavioural potential*, involves the expression of morpho-syntactic categories that are semantically, syntactically and/or pragmatically relevant to the particular function in which a linguistic unit is used, but do not indicate that

⁸ In Croft’s earlier work (1991: 58) *structural coding* is termed “*function indicating morpho-syntax*”.

function as such (cf. Croft 2001: 66). Thus, structural coding can be thought of as a *condition* for using a linguistic (lexical) unit in a certain function, whereas behavioural potential is the *consequence* of using it in that function. Consider for example the Dutch noun *vlees* ‘meat’. If this lexeme is to express the function of modifier in a referential phrase, it needs the affix *-ig*. This affix is an instance of structural coding. However, when the resulting form *vlezig* ‘meaty’ is used to modify a definite neuter noun like *het gezicht* ‘the face’, then it takes the gender agreement suffix *-e* (*het vlezig-e gezicht*). This suffix is an instance of behavioural potential. It is crucial that in Hengeveld’s definitions of PoS classes only structural coding counts as *further morpho-syntactic measures*. In contrast, behavioural potential corresponds to what has been termed ‘morphological criteria’ in section 2.2, and does not play a role in the definitions in (1) above.

Second, it is noteworthy that Hengeveld’s definition of verbs includes a special ‘*only*’-condition, which is absent in the definitions of the other three PoS. This is because in many languages nouns, adjectives, and adverbs can have an additional function next to their defining function as specified in (1) (ii)–(iv): They can also be used as the head of a predicate phrase⁹. This means that verbs are the only PoS class whose exclusive, distinguishing function it is to express the head of a predicate phrase; hence the ‘*only*’-condition in the definition of this lexeme class.

Cross-linguistically, the use of a non-verbal lexeme as the head of a predicate phrase (i.e. non-verbal predication) may trigger various formal expression strategies. Following Hengeveld’s typology of non-verbal predication, two main types of strategies are distinguished: those with a copula and those without one. In the first case, the copula counts as structural coding. The second strategy, without a copula, can be further sub-divided into two types, termed the *zero-1* and the *zero-2 strategy*, respectively. In the case of a zero-1 strategy, the non-verbal predicate shows the same behavioural potential as a verbal predicate in the language under analysis. Thus, any categories that are expressed on a verb, such as TAM and person agreement, should also appear on a non-verbal predicate. In contrast, in the case of a zero-2 strategy the non-verbal predicate does *not* show the same behavioural potential as a verbal predicate, but rather remains uninflected and is simply juxtaposed to its argument¹⁰.

⁹ This is then termed a *non-verbal predication*, which was in fact the main topic of Hengeveld’s 1992 study. See further below.

A third point concerning the details of Hengeveld’s PoS definitions is that the class of adverbs includes *manner* adverbs only. This is because other types of adverbs do not modify the head of a predicate phrase but rather another unit. This may be a larger unit, such as a state of affairs, a proposition, an illocution, or an even larger stretch of discourse¹¹. Alternatively, it may be another modifying unit, such as an adjective or another adverb¹².

Taking into account the issues discussed above, the functions of the four PoS classes (verb, noun, adjective, and manner adverb) can be represented as in Figure 2.3, in which ‘+’ stands for the defining use of a lexical class (as mentioned in the left-hand column); ‘-’ for excluded uses; and ‘+/-’ for potential additional uses. The figure shows that verbs are uniquely used as the head of a predicate phrase, while nouns, adjectives, and manner adverbs may have additional uses as the head of a predicate phrase, next to their respective defining uses as the head of a referential phrase, a modifier in a referential phrase, and a modifier in a predicate phrase.

	Head of Predicate phrase	Head of Referential phrase	Modifier in Referential phrase	Modifier in Predicate phrase
Verbs	+	-	-	-
Nouns	+/-	+	-	-
Adjectives	+/-	-	+	-
Manner adverbs	+/-	-	-	+

Figure 2.3: *The functional potential of parts of speech classes*

2.3.2 The cross-linguistic perspective

2.3.2.1 General introduction

Individual languages across the world do not all have systems of lexical categorization of the type represented in Figures 2.2 and 2.3 above, in

¹⁰ In many cases where a zero-2 strategy is used, a copula can optionally be added. The copula often remains un-expressed in ‘default’ cases, i.e. involving present tense and third person singular. In other, non-default cases, TAM and person distinctions are expressed on the copula. See also Chapter 5, section 5.

¹¹ See Hengeveld & Mackenzie (2008, Chapter 3) for examples.

¹² Regarding this latter point, the modification of lexical units other than verbs, Hengeveld & Van Lier (2008: 760-762) show that the definition of the class of lexemes functioning as a modifier in a predicate phrase must make reference to the PoS class of the head of that predicate phrase. That is to say, when the predicate phrase has a verbal head, it takes different types of adverbial modifiers (namely manner adverbs) than when it has a nominal, adjectival, or adverbial head (namely reference/referent modifiers or degree adverbs).

which there are four PoS classes, each of which is mapped neatly onto a single defining function. Rather, many languages display different types of mappings of their lexical categories onto the set of possible functions. Hengeveld's syntactic approach to lexical categorization provides a method to compare these types of mapping systems across languages, called *parts of speech systems* (henceforth PoS systems).

According to Hengeveld's typology, PoS systems come in two basic types: *rigid* and *flexible* ones. *Rigid* PoS systems are characterized by the fact that all lexical classes are rigid, i.e. specialized for the expression of a single function. When lexical items do appear in other functions, this must be marked by means of further morpho-syntactic measures or structural coding (with the possible exception of the predicative function, cf. the +/- signs in Figure 2.3). Rigid PoS systems may display all four lexeme classes that are represented in Figures 2.2 and 2.3. Alternatively, however, they may lack one or more PoS classes. This means that one or more function(s) cannot be expressed lexically, so that the relevant language must resort to an alternative, non-lexical strategy. In Garo NPs, for instance, the translational equivalent of an English adjective is always expressed by means of a relative clause, as illustrated in (2):

Garo (Hengeveld et al. 2004: 531)¹³

- (2) *da'r-gipa mande*
 be.big-PTC man
 'the big man' (lit. the man who is big/the man who 'big-s')

Thus, in a rigid PoS system all major lexeme classes are functionally specialized for the expression of a single function, even though there may not be a PoS class available for every function.

In contrast, *flexible* PoS systems are characterized by the fact that one or more of the available lexeme classes is flexible, meaning that its members can express *two or more* functions, without further morpho-syntactic measures. In Dutch, for instance, the translational equivalents of English adjectives

¹³ The suffix *-gipa* in this example is not a lexical derivational suffix, as can be seen from example (1), in which the participial form marks a relative clause construction (Burling 2004: 301):

- (1) [*nok-o pisa-ko nik-gipa metra*]
 house-LOC child-ACC see-PTC woman
 'the woman who saw the child at the house'

and manner adverbs belong to a single class of flexible lexemes that can be used without further measures as modifiers in both referential and predicate phrases. This is shown in examples (3) and (4), respectively.

Dutch

(3) [*een mooi meisje*]_{Referential Phrase}
 a beautiful girl
 ‘a beautiful girl.’

(4) *Het meisje* [*zingt mooi*]_{Predicate Phrase}
 The girl sings beautiful
 ‘That girl sings beautifully.’

Notice that in the English translation of (4) the adjective *beautiful* does require a further measure in order for it to be usable in the function of modifier in a predicate phrase, namely the suffix *-ly*¹⁴.

2.3.2.2 The Parts of Speech Hierarchy (Hengeveld 1992)

Hengeveld (1992) shows that the two basic types of PoS systems, flexible and rigid, come in several variants across languages. He claims that this variation can be systematically described in terms of a *Parts of Speech (PoS) Hierarchy*, which is given in (5):

(5) Head of Predicate phrase ⊂ Head of Referential phrase ⊂ Modifier in Referential phrase ⊂ Modifier in Predicate phrase

Proceeding through this hierarchy from left to right, the chance decreases that a language has a separate, functionally specialized lexical class to express that particular function. For a language with a rigid PoS system this means that if it lacks a lexeme class for just a single function, this will be the rightmost function in the hierarchy: modifier in a predicate phrase. In other words, this language will have no class of manner adverbs. In addition, the

¹⁴ In Dutch, modifiers in referential phrases show gender agreement with the head noun in definite contexts. Thus, if in a phrase like (4a) the definite article *het* (rather than indefinite *een*) would be used, then *mooi* would carry the agreement marker *-e*. However, as mentioned above, gender agreement does not count as ‘further measures’, since it constitutes *behavioural potential* rather than *structural coding*.

hierarchy is implicational. As such, it predicts that, if a rigid PoS system lacks a lexeme class for a specific function in the hierarchy, then it will also lack lexeme class(es) for any function(s) further to the right in the hierarchy. For example, a language without a lexeme class to express the function of modifier in a referential phrase will also lack a lexical class for the function of modifier in a predicate phrase. In other words, a language without adjectives is predicted to also lack manner adverbs.

For languages with flexible PoS systems, the hierarchy predicts that they are most likely to have a flexible lexeme class that can express the two rightmost functions: modifier in a referential phrase and modifier in a predicate phrase (as in the Dutch examples (3) and (4) above). Moreover, if in a flexible PoS system a lexeme class can express two particular functions in the hierarchy, it is predicted that this class will also be able to express any function(s) further to the right in the hierarchy, as well as any function(s) lying in between. For example, if a language has a class of lexemes that can be used as the head of a referential phrase and as a modifier in a predicate phrase, then it is predicted that this class will also be used for modification in a referential phrase. In other words, if a language has a class of lexemes that can function as nouns and as manner adverbs, then these lexemes should also be usable as adjectives.

In its original formulation (as given in (5) above), the PoS Hierarchy predicts seven different PoS systems, which are represented in Figure 2.4 below (Hengeveld 1992: 69; Hengeveld et al. 2004: 537). Three PoS systems are flexible, i.e. they include one lexeme class that can be used in multiple functions. The other four systems are rigid, i.e. they involve functionally specialized classes only. The labels for the flexible PoS classes are taken from Hengeveld et al. (2004): A *contentive* is a member of a maximally flexible class that can be used in all four functions; a *non-verb* can be used as the head and modifier in a referential phrase, and as a modifier in a predicate phrase; and a *modifier* is a lexeme that can be used as a modifier in both predicate and referential phrases.

Parts of speech system		Head of Predicate phrase	Head of Referential phrase	Modifier in Referential phrase	Modifier in Predicate phrase
Flexible	1	concentive			
	2	verb	non-verb		
	3	verb	noun	modifier	
Rigid	4	verb	noun	adjective	manner adverb
	5	verb	noun	adjective	–
	6	verb	noun	–	–
	7	verb	–	–	–

Figure 2.4: Possible parts of speech systems according to Hengeveld 1992

2.3.2.3 Intermediate systems (Hengeveld, Rijkhoff & Siewierska 2004)

In more recent work by Hengeveld et al. (2004), the typology in Figure 2.4 is expanded in order to accommodate languages with so-called *intermediate* PoS systems (as first proposed by Smit (2001)). Such intermediate systems exhibit features of two ‘neighbouring’ PoS systems, i.e. two PoS systems that are on contiguous rows in Figure 2.4.

In a flexible intermediate PoS system, the lexeme class that is positioned rightmost in the hierarchy consists of two sub-classes: one containing basic members and the other derived members. The sub-class of derived items can express one function less than the subclass of basic items. In particular, the left-most function that can be expressed by the basic items cannot be expressed by the derived items. For example, Turkish has a PoS system with classes of basic verbs and basic non-verbs. The non-verbs, the right-most lexical class in the system, serve as the input category for derivation with the suffix *-CA*, which creates a class of modifiers, i.e. derived items that can be used as modifiers in referential and predicate phrases, but not as the head of a referential phrase. This is represented in Figure 2.5, in which the shading indicates the derived nature of the modifier class.

Head of Predicate phrase	Head of Referential phrase	Modifier in Referential phrase	Modifier in Predicate phrase
Verbs	Non-verbs		
		Derived modifiers	

Figure 2.5: The flexible intermediate PoS system of Turkish

In (6a-c) the functional possibilities of the Turkish non-verb *güzel* are illustrated: in (6a) it functions as the head of a referential phrase, in (6b) as a modifier in a referential phrase, and in (6c) as a modifier in a predicate phrase.

Turkish (Göksel & Kerslake 2005: 49)

- (6) a. *güzel-im*
 beauty-1POSS
 'my beauty'
- b. *güzel bir köpek*
 beauty ART dog
 'a beautiful dog'
- c. *güzel konuştu*
 beauty s/he.spoke
 'S/he spoke well'

Examples (7a-b) show the distribution of *CA*-derived modifiers: in (7a) *erkek-çe* functions as a modifier in a referential phrase, and in (7b) *ıyı-çe* modifies a predicative head.

Turkish (Kornfilt 1997: 92)

- (7) a. *Erkek-çe bîrses*
 man-MODIF a.voice
 'a manly voice'
- b. *Hasan gömleğ-i ıyı-çe yıka-dı*
 Hasan shirt-ACC good-MODIF wash-PST
 'Hasan washed the shirt well.'

A rigid PoS system is classified as intermediate when its rightmost lexical category is a small, closed class, rather than a large, open one. For example, Alambak has large open classes of rigid verbs and nouns, and a reasonably large class of adjectives (45 items). Its class of manner adverbs, however, comprises only 9 items (Bruce 1984: 87, 88). This is represented in Figure 2.6, in which the darker shading is used to indicate the small, closed nature of the manner adverb class.

Head of Predicate phrase	Head of Referential phrase	Modifier in Referential phrase	Modifier in Predicate phrase
Verbs	Nouns	Adjectives	Manner adverbs

Figure 2.6: *The rigid intermediate PoS system of Alambak*

The case of Alambak point to the lack of a sharp distinction between ‘large, open’ and ‘small, closed’ word classes. The class of adjectives in Alambak is much smaller than the classes of nouns and verbs, but also substantially larger than the class of manner adverbs. Also the notion of openness – the ease with which new members are added to a lexeme class – seems to be of a relative, rather than an absolute nature. As a result, there may be borderline cases, the classification of which inevitably involves a degree of arbitrariness. I will return to this issue in some more detail in Chapter 5 (section 5.4.2.1), where the PoS systems of the individual sample languages are classified.

The intermediate PoS systems predicted by the PoS Hierarchy in (6) are included in Figure 2.7, in which the cells with light shading are derived classes, and darker shaded cells represent small, closed lexical classes (cf. Hengeveld et al. 2004: 409).

Parts of speech system		Head of Predicate phrase	Head of Referential phrase	Modifier in Referential phrase	Modifier in Predicate phrase
Flexible	1	contentive			
	1/2	contentive		non-verb	
		2	verb	non-verb	
	2/3	verb		non-verb	
		3	verb	noun	modifier
	3/4	verb	noun	modifier	manner adverb
Rigid	4	verb	noun	adjective	manner adverb
	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 2.7: Parts of speech systems, including intermediate types

2.3.2.3 More diversity (Hengeveld & Van Lier 2008, 2009)

Even more recent typological investigations make clear that the set of PoS systems predicted by the PoS Hierarchy, including the ‘extended’ set represented in Figure 2.7 above, is too restricted to account for the attested cross-linguistic variation (Hengeveld & Van Lier 2008, 2009). In particular, there are languages that display intermediate PoS systems that are not contiguous in Figure 2.4 above. Moreover, the ‘extra’ classes of these intermediate systems do not necessarily consist of derived members in flexible PoS systems, nor are they always small, closed classes in rigid PoS systems (cf. Figure 2.7 above).

Accommodating these findings, Hengeveld’s earlier model has recently been adapted in order to give it higher coverage, without seriously diminishing its accuracy. In other words, the new model predicts a higher amount of different types of PoS systems, and as such accounts for a larger range of variation attested in the world’s languages, without predicting too many PoS systems that have not (or at least not yet) been attested in any language. Hengeveld and Van Lier (2008, 2009) argue that the four syntactic slots

distinguished in Figure 2.1 above should be re-interpreted as *propositional functions*¹⁵. These functions constitute dimensions of a space defined in terms of two parameters, each of which represents a dominance relation between two functional values. In addition, the two parameters stand in a dominance relation with respect to one another. Together, these hierarchical relations yield a set of predictions concerning possible mappings of PoS categories onto the space of propositional functions. The details of this model are presented in the remainder of the present sub-section.

The first and primary parameter involves the distinction between *predication* and *reference*¹⁶. These two functions are conceived of as the basic communicative functions, with which a speaker builds up the communicative content he or she wishes to express. It is assumed that the function of predication is privileged in relation to the function of reference, since the communicative act of referring presupposes that something is predicated of a referent. For instance, in a noun phrase like ‘*a yellow car*’ the properties ‘*car*’ and ‘*yellow*’ are predicated of the entity being referred to. The hierarchical relation between predication and reference is represented in (8):

(8) Predication \subset Reference

The second and also secondary parameter involves the distinction between *heads* and *modifiers*. Heads are obligatory and therefore primary in relation to optional modifiers¹⁷. This dominance relation is also shown by the fact that the lexical class of a modifier is dependent on the lexical class of its head (see note 12). The hierarchical relation between heads and modifiers is represented in (9):

(9) Head \subset Modifier

¹⁵ Croft (2001) considers modification as a propositional function on a par with predication and reference (see section 2.4.2 of the present chapter). In his earlier work (Croft 1991: 52), however, modification is defined as an “*accessory function to reference and predication*”. This latter definition is in accordance with the treatment of the modifier function in the Hengeveldian model.

¹⁶ Hengeveld and Van Lier (2008) use the term *ascription* where I use *predication*. The former term is used in the framework of *Functional Discourse Grammar* (FDG), which is the follow-up of Functional Grammar (FG), to distinguish between a propositional or *interpersonal* function and a semantic or *representational* function (see Hengeveld & Mackenzie 2008).

¹⁷ cf. Bloomfield’s (1962: 194–196) definition of a *head*: the element of a construction that can function like the whole construction.

In addition, as already mentioned, the two hierarchical relations in (8) and (9) are in turn ranked with respect to one another. In view of the fact that there are appositional languages that do not use modification at all, while there are no languages that do not display the predication-reference distinction (even though they may not do so at the level of lexical classification, but rather at the level of morpho-syntactic constructions), it is expected that the predication-reference parameter is primary in relation to the head-modifier parameter. This is represented in (10):

(10) ((Predication/Reference) \subset (Head/Modifier))

On the basis of these three hierarchical relations, three implicational constraints can be formulated, as listed in (11)-(13):

(11) Predication \subset Reference

- a. If a language has a rigid class of lexemes that can be used as the head of a referential phrase, it must also have a rigid class of lexemes that can be used as the head of a predicate phrase.
- b. If a language has a flexible class of lexemes that can be used as the head of a referential phrase (but not as the head of a predicate phrase, since then the restriction becomes irrelevant), it must also have a flexible or rigid class of lexemes that can be used as the head of a predicate phrase.

(12) Head \subset Modifier

- a. If a language has a rigid class of lexemes that can be used as the modifier within a phrase, it must also have a rigid class of lexemes that can be used as the head of that phrase.
- b. If a language has a flexible class of lexemes that can be used as the modifier within a phrase (but not as the head of that phrase, since then the restriction becomes irrelevant), it must also have a flexible or rigid class of lexemes that can be used as the head of that phrase.

(13) ((Predication/Reference) \subset (Head/Modifier))¹⁸

If a language has distinct (rigid or flexible) classes of lexemes for

heads and modifiers within any phrase, then it must also have distinct (rigid or flexible) classes of lexemes for heads of predicate versus referential phrases.

The parameters in (8)-(10) and the constraints in (11)-(13) are accounted for in the layered implicational map in Figure 2.8 below. This figure shows that the Predication-Reference parameter ranks higher than the Head-Modifier parameter (as indicated by the symbol \sqsubset in the centre), and that the Head-Modifier parameter applies in the domains of both Predication and Reference, which does not exclude the possibility that these domains share a single lexeme class.

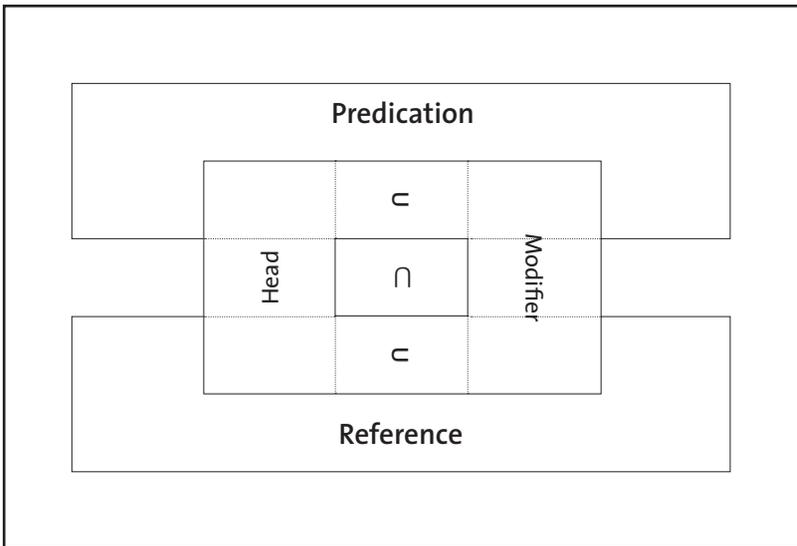


Figure 2.8: The implicational map of parts of speech

The map in Figure 2.8 predicts all seven systems that were also consistent with the original parts of speech hierarchy in (5). These systems are represented in (14)-(20) below. In addition, the new model predicts another 10 systems, which would have been counterexamples to the original model. These are represented in (21)-(30). In what follows, I briefly discuss each of the systems predicted by the map in Figure 2.8. Note that in all representations of PoS

¹⁸ It is explicitly not the case that the Head-Modifier distinction must apply within the predicative domain before it can apply in the referential domain. Thus the combination of (8)/(11) and (9)/(12) should not be read as $(\text{PredHead} \sqsubset \text{PredModifier}) \sqsubset (\text{RefHead} \sqsubset \text{RefModifier})$. Thanks to Michael Cysouw for pointing this out.

systems the flexible PoS classes have dark shading, while rigid classes are in a lighter shade. When there is no PoS class available to express a particular function, then the cell corresponding to this function remains unshaded and is marked with a bar.

The system in (14) is consistent with the new model because it involves no lexical distinctions at all. All functions can be expressed by a single class, called *contentives* (cf. Figures 2.4 and 2.7 above).

(14)

	Head	Modifier
Predication	Contentive	
Reference		

The system in (15) is consistent because (i) it has a flexible lexeme class (*non-verbs*) that can be used as the head of a referential phrase and a rigid lexeme that can be used as the head of a predicate phrase (*verbs*) (cf. (11b)); (ii) it has a flexible class of lexemes that can be used as a modifier within a predicate phrase and a rigid class of lexemes that can be used as the head of that phrase (cf. (12b)); and (iii) it has distinct classes for heads and modifiers in the predicate phrase as well as distinct classes for heads of predicative versus referential phrases (cf. (13)).

(15)

	Head	Modifier
Predication	Verb	
Reference	Non-verb	

The system in (16) is consistent because it involves a predication-reference distinction for heads (*verbs* and *nouns*) but not for *modifiers*. This is in accordance with constraints (11a), (12b) and (13).

(16)

	Head	Modifier
Predication	Verb	Modifier
Reference	Noun	

The system in (17) is consistent with the model because it involves head-modifier distinctions in both the predicative and the referential domain. Thus it has *verbs*, *nouns*, *adjectives*, and *manner adverbs*.

(17)

	Head	Modifier
Predication	Verb	Manner adverb
Reference	Noun	Adjective

The system in (18) differs from the one in (17) only in that it lacks a lexeme class for the function of modification in predicate phrases (*manner adverbs*), while the head function in this domain is expressed by *verbs*. This does not violate any of the constraints.

(18)

	Head	Modifier
Predication	Verb	–
Reference	Noun	Adjective

The system in (19) again involves a predication-reference distinction in the head domain, but does not have lexical classes for either of the two modifier functions. This system conforms to (11a), since it has rigid classes of heads in both the reference and the predication domain (*verbs* and *nouns*), while constraints (12) and (13) are not applicable.

(19)

	Head	Modifier
Predication	Verb	–
Reference	Noun	–

The system in (20), finally, has only one lexical class, namely one that fulfils the function of head of a predicate phrase (*verbs*). This does not violate any of the constraints.

(20)

	Head	Modifier
Predication	Verb	–
Reference	–	–

These were the PoS systems predicted by both the original PoS hierarchy in (5) and the implicational map in Figure 2.8. As mentioned, the latter model predicts another 10 possible PoS systems, which would have presented counterexamples to the former model. These systems are represented in (21)-(29) below.

The system in (21) is in accordance with constraint (11b), since it has a class of flexible lexemes that can be used as referential heads (*nominals*) and a rigid class for predicative heads (*verbs*). It is also in accordance with (12a) in the predication domain, where it has rigid modifiers (*manner adverbs*) and rigid heads. Finally, it conforms to (13), since it has distinct classes for heads and modifiers in the predicate phrase, as well as distinct classes for heads in predicative versus referential phrases.

(21)

	Head	Modifier
Predication	Verb	Manner adverb
Reference	Nominal	

The system in (22) is in accordance with constraint (11b), since it has a class of flexible lexemes that can be used as the head of a referential phrase (*nominals*), and a class of rigid lexemes for the function of head of a predicate phrase (*verbs*). Constraint (12) and (13) are not applicable.

(22)

	Head	Modifier
Predication	Verb	–
Reference	Nominal	

The system in (23) involves a lexical distinction between predication and reference, but not between heads and modifiers. The PoS classes in this

system are termed *predicatives* and *nominals* (see also Hengeveld & Van Lier 2009). The system is in accordance with (11b). Constraints (12) and (13) are not applicable.

(23)

	Head	Modifier
Predication	Predicative	
Reference	Nominal	

The system in (24) conforms to constraint (11b), since it has a flexible class of lexemes that can be used predicatively (*predicatives*), while not having a class of lexemes that can be used referentially. Constraints (12) and (13) do not apply.

(24)

	Head	Modifier
Predication	Predicative	
Reference	–	–

The system in (25) involves a single flexible class of *heads*, i.e. lexemes that can be used as the head of both predicative and referential phrases. This system does not violate any of the constraints.

(25)

	Head	Modifier
Predication	Head	–
Reference		–

The system in (26) also involves a single flexible class, but one that can be used in all functions except modifier in a predicate phrase. This system does not violate any of the constraints¹⁹. For want of a good candidate, no special term

¹⁹ Note that the *lack* of a lexeme class for the function of modifier in a predicate phrase does not mean that there is a lexical head-modifier distinction within the functional domain of predication.

is proposed for this type of PoS class. It is simply indicated as ‘flexible class A’ (*Flex A*); its functional distribution can be read off the representation.

(26)

	Head	Modifier
Predication		–
Reference	Flex A	

The system in (27) is similar to the one in (26), except that the flexible class can now be used in all functions except modifier in a referential phrase. Again, no special term is introduced for this type of flexible lexeme class; it is indicated as *Flex B*.

(27)

	Head	Modifier
Predication		
Reference	Flex B	–

The system in (28) involves a rigid class of verbs and a flexible class of lexemes, termed *Flex C*, which can be used as the head of a referential phrase and a modifier in a predicate phrase, but not as a modifier in a referential phrase. As such, it is in accordance with constraint (11b), with (12b) in the predication domain, and with (13).

(28)

	Head	Modifier
Predication	Verb	Flex C
Reference	Flex C	–

The system in (29) involves two ‘cross-wise’ flexible lexeme classes: One class (*Flex C*; see above) that can be used as the head of a referential phrase and as a modifier in a predicate phrase, while the other class, termed *Flex D*, can be used as the head of a predicate phrase and as a modifier in a referential phrase. This system does not violate any constraint.

(29)

	Head	Modifier
Predication	Flex D	Flex C
Reference	Flex C	Flex D

Finally, the system in (30) has rigid lexeme classes for heads and modifiers in the predication domain (*verbs* and *manner adverbs*), while in the reference domain it has a rigid lexeme class for heads only (*nouns*). This system is in accordance with constraints (11a), (12a) and (13).

(30)

	Head	Modifier
Predication	Verb	Manner adverb
Reference	Noun	–

In sum, the implicational map model in Figure 2.8 predicts 17 possible PoS systems (of which 7 were also predicted by the original PoS hierarchy), which are expected to occur in actual languages. In addition however, the model predicts that a large number of logically possible PoS systems will *not* occur in any language. In particular, 34 out of 51 logically possible systems are excluded by the model, as they violate one or more of the constraints in (11), (12) and (13) above. For expositional reasons, these excluded systems are listed separately in Appendix 11, where I specify for each system which constraint(s) it violates. In Chapter 5 the complete set of predicted and excluded PoS systems will be compared with the PoS systems that are (un-) attested in the sample languages.

At this point, it should be noted that the PoS systems presented above do not include any intermediate systems. It is expected, however, that the predicted systems can indeed be supplemented with additional PoS classes, as long as this does not violate the constraints in (11)-(13). As I mentioned earlier, these 'extra' classes need not be contiguous to the rightmost large basic class of a system, in terms of the original PoS hierarchy. Moreover, and again contrary to the predictions of the original hierarchy, extra classes are not necessarily derived, small and/or closed. Rather, they can also be large, open classes consisting of simple members. This will be illustrated in full

detail in Chapter 5. However, to give a flavour of the attested cross-linguistic variation in PoS systems, I provide some examples below.

First, languages with flexible PoS systems may have extra classes that are large and open rather than derived. Santali, for instance, combines a maximally flexible PoS class of contentives with a large, open class of rigid verbs (Neukom 2001: 17). This is represented in Figure 2.9. Note that this system would have been a counterexample to the original theory, which predicts that an ‘intermediate’ system with contentives can only involve a class of derived non-verbs (cf. Figure 2.7).

	Head	Modifier
Predication	Derived nouns	Contentives
Reference		

Figure 2.9: The PoS system of Santali

Second, consider Garo, a language with a rigid PoS system. It has large open classes of verbs and nouns, no adjectives, a small class of simple manner adverbs, and a large class of verb-derived manner adverbs (Burling 2004). This is represented in Figure 2.10. The PoS system of Garo would have been a counterexample to the original PoS hierarchy, since there is no lexical class available for the function that is positioned between the functions of modifier in a predicate phrase and head of a referential phrase. Garo’s PoS system does not fit into the extended typology of PoS systems in Figure 2.7 either, since the latter does not take into account the possibility of derived classes in rigid systems.

	Head	Modifier
Predication	Verbs	Derived/small class of simple manner adverbs
Reference	Nouns	–

Figure 2.10: The PoS system of Garo

These examples give a glimpse of the diversity encountered in the realm of PoS systems across languages, showing that the range of variation is much wider than Hengeveld’s original typology suggests.

2.3.3 Summary

In this section I have presented Hengeveld's theory of lexical categorization, from its original formulation to its most recent version. I explained that the Hengeveldian approach has two crucial characteristics. The first one is the exclusion of semantics as a criterion for the definition of PoS. The second important component involves the 'further morpho-syntactic measures' that do or do not accompany the use of a particular group of lexemes in a particular functional environment. Such further measures include structural coding only; behavioural potential, or morphological criteria, are not taken into account.

Hengeveld's approach avoids (but does not provide a solution to) the problems associated with semantic and morphological criteria for PoS definitions, as pointed out in section 2.2. In particular, the Hengeveldian method allows for a straightforward cross-linguistic comparison of the various ways in which languages may map groups of lexemes onto a set of propositional functions. However, this convenient situation comes at a cost. In the next section I discuss the problems associated with Hengeveld's method, and the alternatives that have been proposed for it.

2.4 Critique on the syntactic approach

2.4.1 Introduction

The most elaborate and critical alternative to Hengeveld's approach is Croft's *Universal-Typological Theory of Parts of Speech* (2000, 2001). The crucial difference between the two approaches is that Croft's theory, unlike Hengeveld's, takes into account semantic and morphological criteria. According to Croft, leaving out these two types of linguistic data implies missing out on the cross-linguistic generalizations that apply to them.

First, in section 2.4.2, I start with a brief outline of Croft's theory. Subsequently, in section 2.4.3, I discuss in more detail the specific objections that Croft has raised against Hengeveld's approach. In fact, these objections are not confined to Hengeveld's PoS theory. Rather, they target any theory of linguistic categories that fails to take into account the full range of linguistic evidence, and that fails to acknowledge that such a complete distributional analysis inevitably leads to the conclusion that cross-linguistically identical categories do not exist²⁰.

²⁰ For a recent general assessment of this fundamental issue in linguistic typology, the reader is referred to Haspelmath (2007) and Newmeyer (2007).

Finally, one particular aspect of Croft’s critique on Hengeveld’s method concerns the role of semantics in languages with flexible word classes. Since this is a complex issue that has been the subject of a rather extensive discussion in recent literature, I treat it separately in section 2.5.

2.4.2 Croft’s Universal-Typological Theory of Parts of Speech

Croft (2000, 2001) distinguishes three functions that lexemes may express: predication, reference, and modification. These are termed *propositional act functions*. Although this resembles the Hengeveldian model, in the latter the modifier function is treated as a dimension of the head-modifier distinction, which cross-cuts the predication-reference distinction, rather than as a basic function in itself (cf. note 15).

A much more important difference between the two theories is that Croft divides the lexemes that may fulfil the three propositional act functions into three different *semantic classes*: objects, properties, and actions. These semantic classes are in turn defined in terms of the following four semantic properties (Croft 2001: 87):

- (i) *Relationality*: whether a definition of a concept inherently requires reference to another concept;
- (ii) *Stativity*: whether the concept represents a state or a process;
- (iii) *Transitoriness*: whether the concept represents a transitory state or process or an inherent or permanent state of the entity in question;
- (iv) *Gradability*: whether the entity is graded along a scalar dimension.

The definitions of the semantic classes in terms of these properties are presented in Figure 2.11 below (taken from Croft 2001: 87)²¹.

	relationality	stativity	transitoriness	gradability
Objects	non-relational	state	permanent	non-gradable
Properties	relational	state	permanent/ transitory	gradable
Actions	relational	process	transitory	non-gradable

Figure 2.11: Properties of semantic classes

In principle, lexemes belonging to any of the three semantic classes (objects, properties, and actions) may express any of the three propositional act functions (reference, modification, and predication). However, the

assumption of the prototype approach (first formulated in Croft 1991) is that the semantic classes of objects, properties, and actions are the *typological prototypes* of referring, modifying, and predicating constructions, respectively. Croft thus defines PoS as prototypical or *unmarked* combinations of semantic class and propositional function. A prototypical noun denotes an object and functions referentially; a prototypical adjective denotes a property and functions as a modifier; and a prototypical verb denotes an action and functions predicatively. This is represented in Figure 2.12 (adapted from Table 2.3. in Croft 2001: 88). Any combination of propositional act function and semantic class other than the three prototypical ones represented in Figure 2.12 is claimed to be typologically *marked*.

	reference	modification	predication
Objects	Unmarked nouns		
Properties		Unmarked adjectives	
Actions			Unmarked verbs

Figure 2.12: Nouns, adjectives and verbs as prototypical combinations of propositional act functions and semantic classes

Importantly, although Croft defines PoS as typological prototypes, rather than as language-specific categories, this by no means implies that he regards PoS-typology as a pointless enterprise. On the contrary, the conceptual notion of typological markedness (versus unmarkedness or prototypicality), as it applies to the combinations of propositional functions and semantic categories, correlates cross-linguistically with formal phenomena. In

²¹ In an earlier version of his theory (1991), Croft uses different labels for some of the semantic properties listed in Figure 2.11. Pustet (2000, 2003) draws heavily on Croft's work, but she proposes a slightly different semantic system. Apart from the property of *transitivity/valency*, which is the same as Croft's (1991), she introduces a new semantic parameter, namely *energy release* (renamed *dynamicity* in Pustet 2003) and leaves out Croft's parameter of *gradability*. The latter is apparently integrated, together with *perfectivity* and *processuality* (corresponding to Croft's *stativity* and *persistence/transitoriness*), into the notion of *transcience*. Just for the sake of clarity, this rather confusing terminological situation is represented in the schema below:

Croft 1991	Croft 2001	Pustet 2000	Pustet 2003
valency	relationality	transitivity	transitivity/valence
–	–	energy release	dynamicity
persistence	transitoriness	perfectivity	–
stativity	stativity	processuality	transcience
gradability	gradability	–	–

particular, Croft identifies two implicational universals, which constrain cross-linguistic structural variation in terms of both structural coding and behavioural potential.

The first universal, pertaining to structural coding, predicts that a typologically *marked* combination of pragmatic function and semantic class must be encoded by at least as many morphemes as an *unmarked* combination. The second universal makes the opposite prediction for behavioural potential: It states that the range of behavioural potential displayed in the case of an *unmarked* combination of pragmatic function and semantic class must be at least as wide as the range displayed in the case of a *marked* combination (Croft 2001: 90, 91). Notably, both universals allow for the possibility of a typologically marked combination showing exactly the same amount of morpho-syntactic measures as an unmarked combination, including zero-marking²². What is excluded is a situation in which a marked combination shows less structural coding or more behavioural potential than an unmarked combination²³.

In sum, Croft's PoS theory defines a *universal conceptual space* that is made up of a pragmatic (propositional) and a semantic (denotational) dimension. Onto this conceptual space, a set of language-specific lexical categories is mapped, which are identified on the basis of a distributional analysis of formal criteria, including both structural coding and behavioural potential. Croft's theory and Hengeveld's theory thus resemble one another to the extent that both are concerned with form-function mappings (or 'construction-filler interactions', as Croft (2005: 437) calls them). However, their approaches differ in two fundamental respects. First, Croft's functional space includes a semantic dimension, which is lacking in the Hengeveldian model. Second, Croft is interested not only in structural coding – as Hengeveld is in further morpho-syntactic measures –, but also takes into account all other distributional properties of lexemes, i.e. their behavioural potential.

According to Croft, the theoretical repercussions of using only a subset of morpho-syntactic criteria are serious. He argues that constructions or

²² As far as structural coding is concerned, this would amount to Hengeveld's definition of flexibility.

²³ Note that the amount of marking is measured in terms of the number of morphemes (Croft 2001: 90, 91). This is not a very straightforward method, though, since it is not clear how to treat for instance portmanteau morphemes. Therefore, it would probably be better to count the number of semantic categories expressed on the lexeme, either by bound (agglutinative or fusional) morphemes or by analytic grammatical particles.

categories (including PoS classes) identified in this way are “*epiphenomenal*” and that “*one will not find language universals that way*” (Croft 2005: 436). In what follows, I discuss these claims in some more detail.

2.4.3 Lexical categories within and across languages:

The generality and the subclass problems

Croft’s main objection to Hengeveld’s PoS theory is that it is selective about which distributional tests are taken into account (structural coding) and which are not (behavioural potential). According to Croft, Hengeveld’s approach amounts to “*methodological opportunism*”, which can be avoided only by taking into account the complete set of distributional facts in each individual language. This, he argues, will reveal “*covert categories, that is, categories not obligatorily flagged by overt structural coding*” (Croft 2001: 75; 2005: 434)²⁴.

However, the approach advocated by Croft raises two potential problems, which are in close relation to each other, and which have been termed the *subclass problem* and the *generality problem* (Haspelmath 2001). The former problem is concerned with intra-linguistic lexical categorization; the latter with inter-linguistic comparability of lexical categories.

Starting with the subclass problem, if one takes into account the full set of distributional facts about groups of lexemes in a specific language, this gives rise to a “*myriad*” of different types of sub-categorizations, without any means to decide which of these categories are to be considered ‘PoS classes’ and which ones are not (Croft 2001: 78ff, Haspelmath 2001: 16540). This problem is also aptly described by Schachter and Shopen:

“It must be acknowledged [...] that there is not always a clear basis for deciding whether two distinguishable open classes of words that occur in a language should be identified as different parts of speech or as subclasses of a single part of speech. The reason for this is that the open parts of speech classes must be distinguished from one another on the basis of a *cluster* of properties [...]. Typically there is some overlap, some sharing of features, as well as some differentiation. [...] What this means is that there may in some cases be considerable arbitrariness in the identification of two open word classes as distinct parts of speech

²⁴ Croft (2001: 73) also points out that a PoS theory like Hengeveld’s, which lacks a semantic dimension, is bound to overlook regularities in the relevant domain, such as those applicable to semantic shifts in flexible lexemes. This issue will be discussed further in section 2.5.

rather than subclasses of a single part of speech.” (Schachter & Shopen 2007: 4; emphasis in the original)

Croft (2001: 81-83, following Uehara 1998) illustrates this point with an example from Japanese, which I will discuss here in some detail. In Japanese, the boundary between nouns and adjectives is not clear-cut. Traditionally, three categories are distinguished: ‘nouns’, ‘adjectives’, and ‘nominal adjectives’. ‘Nouns’ can function without structural coding as the head of a referential phrase. When they are used as the head of a predicate phrase they need a copula, and when they are used as a modifier in a referential phrase they take a genitive particle *no*. ‘Nominal adjectives’ behave in the same way as nouns when functioning as the head of a referential phrase or the head of a predicate phrase, but they take the linking particle *na* when used in modifier function. ‘Adjectives’ can not be used as heads in referential phrases. In contrast, they can be used without structural coding as referential modifiers and as heads of predicate phrases. In both functions, the behavioural potential of adjectives consists of tense-marking: *-i* for non-past and *-katta* for past (see Hinds 1986: 44)²⁵.

However, a number of other lexical classes may be distinguished in Japanese with distributional patterns that seem to fall somewhere in between the patterns of ‘nouns’, ‘nominal adjectives’, and ‘adjectives’. First, there is a class of lexemes that sometimes behave like nouns and sometimes like nominal adjectives. In particular, when functioning as modifiers, these items can appear either with the genitive marker *no*, like nouns (see (31a)), or with the linker *na*, like nominal adjectives (see (31b)):

Japanese (Croft 2001: 82)

- (31) a. *heiwa no sisyā*
peace GEN messenger
‘a messenger of peace’
- b. *heiwa na kuni*
peace(ful) LK country
‘a peaceful country’

²⁵ In Chapter 5 I will interpret the fact that ‘adjectives’ inflect for tense when used as modifiers as evidence that these lexemes are in fact verbs that may be used as dependent predicates in (zero-marked) relative clauses. At present, however, I will follow the argument as presented by Croft.

The members of another such ‘in-between class’ behave partly like nominal adjectives and partly like adjectives. One part occurs either with a copula when used as the head of a predicate phrase, or without a copula and with tense-marking (see (32a)). In modifier function these items can take either the linker *na* or zero structural coding with tense inflection (see (32b)).

- (32) a. *ataka-i* / *ataka da*
 warm-PRS / warm COP
 ‘It is warm.’
- b. *ataka-i hi* / *ataka na hi*
 warm-PRS day / warm LK day
 ‘a warm day’

The second part of this subclass displays alternative coding strategies only in the modifier function (*na* versus zero structural coding and tense marking, as in (33a), cf. (32b)). When used as the head of a predicate phrase the relevant lexemes always occur without a copula and with tense marking (see (33b) and cf. (32a)):

- (33) a. *tiisa-i hon* / *tiisa na hon*
 small-PRS book / small LK book
 ‘a small book’
- b. *tiisa-i* / **tiisa da*
 small-PRS / small COP
 ‘It is small.’

These distributional facts are summarized in Figure 2.13:

PoS class	Function/Coding		
	Head of Pred Phrase	Head of Ref Phrase	Mod in Ref Phrase
Noun	<i>da</i> (COP)	∅	<i>no</i> (GEN)
Noun/nominal adjective	<i>da</i> (COP)	∅	<i>no</i> (GEN)/ <i>na</i> (LK)
Nominal adjective	<i>da</i> (COP)	∅	<i>na</i> (LK)
Nominal adjective/adjective 1	<i>da</i> (COP)/∅ + Tense	∅	<i>na</i> (LK)/∅ + Tense
Nominal adjective/adjective 2	∅ + Tense	∅	<i>na</i> (LK)/∅ + Tense
Adjective	∅ + Tense	X	∅ + Tense

Figure 2.13: *The noun–adjective distinction in Japanese*

The crucial question is now whether one should interpret all these different distributional patterns as evidence for as many PoS classes in Japanese. If so, then a Hengeveldian typology of PoS systems would be split up into a potentially unlimited number of sub-classifications. For Croft’s theory, however, this problem is irrelevant: From his point of view, it is simply not necessary to decide whether the classes identified above are nouns or adjectives, or both, or neither. They just represent a set of language-specific formal categories mapped onto a universal functional space.

A direct consequence of the subclass problem is the generality problem: lexical categories like the Japanese ones identified in Figure 2.13 are not comparable with lexical categories attested in any other individual language. Since languages differ in terms of the number and type of relevant distributional criteria, a lexeme class that is defined on the basis of such criteria in one language can never be identical to a lexeme class in another language, which expresses different morpho-syntactic distinctions. Again, this is a non-issue in terms of Croft’s prototype approach. The semantic dimension of his theory accounts for cross-linguistic generalizations about the ways in which individual languages cut up the universal functional space cake: the less prototypical the combination of semantic meaning and propositional function, the more structural coding will be required, and the less behavioural potential will be displayed.

The Hengeveldian approach, in contrast, ignores fuzzy boundaries between lexeme classes, such as between nouns and adjectives in Japanese. According to this approach, the relevant facts for Japanese are that it has

(i) a class of lexemes that can be used without structural coding as the head of a referential phrase (i.e. Hengeveld's definition of nouns), and (ii) a class of lexemes that can be used without structural coding as a modifier in a referential phrase (i.e. Hengeveld's definition of adjectives). Nouns and adjectives (and other PoS) defined along these lines are not universal in the way that Croft's prototypes are universal, and they are not language-specific in the Croftian sense either (i.e. they are not based on a complete distributional analysis of a the particular language system of Japanese). However, these PoS are cross-linguistically comparable, even though they can be considered identical across languages only as far as Hengeveld's narrow (or should one say 'broad') definitions go. Adding more analytical primitives, as Croft does, inevitably leads to more categorial differentiation.

2.4.4 Summary, outlook

In sum, Croft's method of identifying lexical categories gives rise to the subclass problem and the generality problem, which concern the intra-linguistic definition and inter-linguistic comparability of PoS classes. However, Croft avoids these problems by defining PoS as prototypes rather than as language-specific categories, and by identifying cross-linguistic generalizations about patterns of relative formal markedness, which make reference to both structural coding and behavioural potential. In contrast, Hengeveld's approach to lexical categorization avoids the subclass and generality problems by looking at a restricted set of distributional data, namely at the presence versus absence of structural coding accompanying the use of particular groups of lexemes in particular propositional functions.

In section 2.5 I turn to another problematic issue that has been raised in relation to Hengeveld's PoS theory. This issue concerns the absence of a semantic dimension in relation to the problem of defining *lexical flexibility*. In particular, since the Hengeveldian theory does not take into account the semantic denotation of lexemes, it has nothing to say about the semantic interpretation of supposedly flexible lexemes in the various functions that these items may express. However, claims about the existence of lexical flexibility (in general as well as with reference to particular languages) have often been challenged on semantic grounds: Even though presumed flexible lexemes do not change formally when used in various functions, they may have different (if related) meanings in each of those functions. Scholars disagree about the origin and theoretical significance of such semantic shifts. In addition to the problem of semantic interpretation, a number of other

issues have entered the discussion about lexical flexibility. The following section reviews this discussion.

2.5 Defining lexical flexibility

2.5.1 Introduction

This section addresses the debate surrounding so-called *flexible languages*. As a point of departure for the discussion, I use the three criteria for lexical flexibility proposed by Evans and Osada (2005). The first and most complicated criterion, to be discussed in section 2.5.2, is concerned with *semantic (non-)compositionality*, i.e. with the semantic interpretation of presumably flexible lexemes in different functional environments. In section 2.5.3 I take up the criterion of *exhaustiveness*, which relates to the problem of quantitative measurement of lexical flexibility. Finally, in section 2.5.4, I discuss the criterion of *equivalent combinatorics*, which pertains to the issue of bi-directional lexical flexibility.

2.5.2 The criterion of compositionality: flexibility and semantic shift

2.5.2.1 Evans and Osada's proposal and some alternatives

Presumably flexible lexemes, i.e. lexemes that can be used without further measures in two or more propositional functions, receive slightly different semantic interpretations, depending on the particular function in which they appear. Although these meanings are (closely) related, particular meaning shifts may differ in terms of their degree of regularity: while some are apparently very predictable, others seem to be rather more idiosyncratic.

Evans and Osada's (2005) *criterion of semantic compositionality* states that, for a lexeme to be 'truly' flexible,

“[A]ny semantic differences between the uses of a putative ‘fluid’ [i.e. ‘flexible’, EvL] lexeme in two syntactic positions must be attributable to the function of that position.” (Evans & Osada 2005: 367)

A corollary of this criterion is that the meaning shift, or what Evans and Osada call the *coercion effect*, should be the same for any member of a flexible lexeme class that is used in a particular function, apart from possible semantic interactions attributable to behavioural potential that belongs to the function, such as for instance aspect markers. According to Evans and

Osada, if the compositionality criterion is not satisfied, i.e. if the semantic interpretation of a flexible lexeme in different functional environments is not entirely compositional, then what looks like flexibility should in fact be analyzed as ‘*rampanť*’ zero-conversion: large-scale derivation from one lexical category to another without any formal reflection.

Similarly, Croft (2000, 2001, 2005) argues that cases of non-compositional semantic shifts are not instances of flexibility. Rather, he argues, they are instances of *polysemy*, i.e. they involve two or more related meanings associated with a single form. Since the relation between these various meanings is not predictable, each of them must be stored individually in the lexicon, and must be labelled as appropriate in a particular functional environment. This amounts to the same kind of categorization that is found in languages with rigid PoS systems.

As we have seen, Hengeveld’s original PoS theory makes no reference to semantics, and therefore it does not account for meaning shifts in flexible lexemes. In Croft’s words: “*Hengeveld ignores what happens to a lexical root’s meaning when used in more than one function*” (Croft 2000: 69). This point is illustrated with examples from Tongan, which were adduced by Hengeveld (1992: 66) as evidence for analyzing Tongan as a flexible language. According to Hengeveld, the examples in (34) demonstrate that the lexeme *si’i* is a flexible lexeme, because it can be used without further measures as the head of a predicate phrase (34a), the head of a referential phrase (34b), and a modifier in a referential phrase (34c)²⁶. In contrast, Croft argues that this analysis is problematic, since it does not account for the different meanings of *si’i* that are associated with each of its various uses.

- (34) a. [Na’e si’i] ‘ae akó
 PST **small** ABS school:DEF
 ‘The school was small.’
- b. ‘i [‘ene si’i]
 in 3SG.POSS **childhood:DEF**
 ‘in his/her childhood’

²⁶ Note in passing that *ako* presumably also represents an instance of lexical flexibility.

- c. *Na'e ako [a tamasi'i si'i] iate au*
 PST study ABS child small LOC 1SG
 'The small child studied at my house.'

In response to Croft's critique, Hengeveld and his colleagues argue that there is no need to assume multiple meanings for flexible lexemes like *si'i* (Hengeveld et al. 2004, Hengeveld & Rijkhoff 2005). In other words, they regard flexible lexemes as *monosemous* rather than polysemous. Meaning shifts, they argue, come about because the use of a flexible lexeme in a particular function activates certain aspects of the lexeme's general meaning. This view is clearly stated in the following quotation:

"[E]ach flexible lexeme has a single (vague) sense. By placing the flexible lexeme in a particular slot or by providing it with certain morphological markers²⁷, the speaker highlights those meaning components of the flexible lexemes 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 the meaning, giving the category neutral lexeme a particular categorical (verbal, nominal, etc.) flavour." (Hengeveld et al. 2004: 541)

However, Croft (2001: 71, 72) argues that problems with this vagueness approach arise in defining the meaning of a lexeme in such a way that it is neutral with respect to the various interpretations associated with it in each of its possible functions. He thus claims that it is impossible to come up with a definition that covers all and only the meanings of a flexible lexeme like Tongan *si'i*. A definition such as 'some concept associated with smallness' is too general because it may include all kinds of concepts that have the property of 'smallness'. As mentioned above, the only alternative, according to Croft, is to specify by convention two or more different meanings of the lexeme in question, and to store these meanings together with the function to which they belong. This analysis undermines the very notion of flexibility.

²⁷ The term 'morphological markers' refers to *behavioural potential*, not to *structural coding*; the presence of the latter would obviously block the analysis as a flexible lexeme. (Kees Hengeveld, p.c.)

Considering the Tongan data in (34) in light of Evans and Osada's criterion of semantic compositionality, one may argue that the shift in the meaning of *si'i* in (34a) ('be small') versus (34c) ('small') is indeed attributable to the different functional environments in which the lexeme appears, namely predication and attribution, respectively. In contrast, this type of analysis cannot easily be applied to (34b), where *si'i* means 'childhood' (or 'period in one's life during which one is small'). This meaning does not follow predictably from the fact that *si'i* functions as the head of referential phrase in this example. Meanings such as 'smallness' or 'small one' appear to be at least as plausible in this function. This would suggest that Tongan does not meet the compositionality criterion, at least not across the board.

However, Hengeveld and Rijkhoff (2005) also disagree with Evans and Osada's assumption that a functional environment *adds* semantic content to a flexible lexeme that it did not have before it was used. As mentioned, Hengeveld and his colleagues claim that flexible lexemes are semantically *vague* (cf. the above quotation). This means that both the verbal and the nominal senses are already contained in the semantics, and that the functional slot "*only highlights meaning components that are already there (...)*" (Hengeveld & Rijkhoff 2005: 514). Thus, according to Hengeveld and Rijkhoff, non-compositional meaning shifts are not an argument *against* flexibility. To the contrary, they say that semantic idiosyncrasies should be expected, since each lexeme has its own unique set of meaning components, which in combination with a particular functional environment yields the intended semantic interpretation.

This line of thought links up with work in the framework of Cognitive Grammar, as developed by Langacker (1987). As mentioned briefly in section 2.2, in Cognitive Grammar the meanings of lexemes are defined as image-schematic concepts. Verb meanings are conceptualized as temporal sequences and are therefore *sequentially* scanned. Nouns, on the other hand, are conceptualized as bounded regions. If the meaning of a noun includes the notion of some temporal unfolding (like a verb), then it is scanned *summarily*, as a whole, rather than *sequentially*. Within this cognitive framework, flexible lexemes can be defined as having a single conceptual structure that is compatible with either a verbal or a nominal meaning. Depending on the functional frame in which the lexeme is used, a profiling of the concept is triggered that corresponds to the appropriate semantic interpretation of the word.

Farrell (2001) applies this cognitive approach to English lexemes. Even though English is probably most often thought of as a language with rigid PoS classes, it has long been noted that many lexemes in this language can indeed be used as nouns and verbs without any overt derivation (Clark & Clark 1979)²⁸. Farrell argues that any English lexeme – irrespective of the behaviour of other lexemes in the language – that can occur in two different functions without derivational marking does not belong to a part of speech until it is inserted into a syntactic frame. In other words, he claims that at least some proportion of the English lexicon is flexible.

Consider for instance the lexeme *kiss*. Farrell analyzes this item as having a conceptual structure that defines contact between the lips of one entity, the agent, with another entity, the patient. This is shown in the upper box of Figure 2.14 (taken from Farrell (2001: 114, Figure 2)). When *kiss* is used as the head of a predicate phrase, its action-interpretation is profiled, as represented in the lower left box of Figure 2.14 by the bold lines of the temporal t-axis and the oval and circle designating the participants. If, on the other hand, *kiss* is used as the head of a referential phrase, then the abstract region defining the overall event is profiled, as represented in the lower right box by the bold print of the outermost oval.

According to Farrell, *kiss* is an example of a so-called *process-centred* word. There are also *thing-centred* words. An example of the latter would be *hammer*. The analysis of thing-centred words differs to the extent that using the lexeme in referential function triggers the profiling of one of the entities involved in the conceptual representation – the *hammer* in this case –, rather than the overall process, as in the case of *kiss*. As Farrell himself admits, however, there are many other possible types of semantic shifts, which do not correspond to the two types just discussed.

In fact, this seems to be exactly what also constitutes the flaw in Hengeveld and Rijkhoff's analysis of flexibility: Even if both the lexeme's particular semantic features and its functional environment are taken into account, it is not always predictable exactly which meaning elements are to be profiled in a particular function. On the other hand, an advantage of Hengeveld and Rijkhoff's vagueness analysis and Farrell's cognitive analysis is that neither of them needs to posit a process of *conversion*, that is,

²⁸ See also Evans & Osada (2005) (and the discussion further on in this section), who analysed samples of 120 lexemes in Mundari (a presumably flexible language) and English, and found that around 75% of the Mundari lexemes could function as verbs and nouns, against around 65% of the English items.

zero-derivation from one lexical category to a new one, without any formal evidence for such a transition. In other words, under both analyses the lack of overt morphological marking of a category change is explained by the fact that there is no such change going on in the first place.

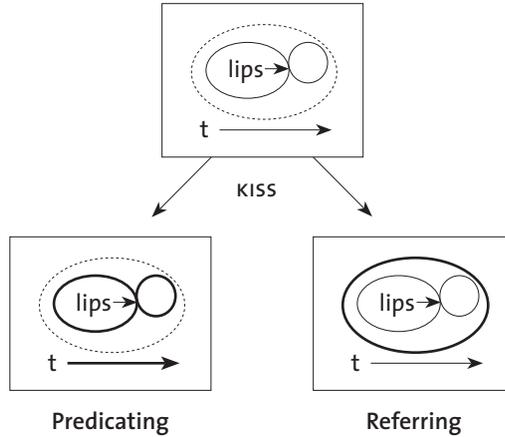


Figure 2.14: The conceptual structure of *kiss*.

In general, the above discussion makes clear that some types of meaning shifts accompanying the use of formally identical lexemes in different functional environments appear to be highly regular, while others seem to be less predictable. In addition, it has been shown that there are, broadly speaking, two views concerning the relation between lexical flexibility and semantic shift: One claiming that ‘true’ flexibility involves purely compositional semantics, and another claiming that idiosyncratic interpretations are expected on the basis of the specific meaning components that are part of the unique semantics of any flexible lexeme. In the next section, I will show that in fact both regular and irregular types of semantic interpretations occur in languages for which pervasive lexical flexibility has been claimed. It will be argued that the relevant data can be understood in terms of a mismatch between lexical and phrase-structural categorization in these languages.

2.5.2.2 An integrative approach to semantics in flexible languages

Theories about PoS often take for granted that there is a one-to-one correlation between lexical categories and syntactic categories. However, Sasse (1993a, b) explicitly advocates a clear distinction between these two types or levels of categorization. Following Sasse’s view, Himmelmann characterizes the levels as follows:

“One level is the level of terminal syntactic categories where lexical items are categorized according to their phrase-structural properties. (...). The second level is the level of lexical categories proper where lexical items are categorized according to those grammatical features which are not directly relevant for phrase structure [...] (i.e. phonological, morphological, and syntactic features).” (Himmelman 2007: 263)

Making this distinction implies that there may or may not be a correlation between the categories distinguished at one level and those distinguished at the other level. The former situation, i.e. a one-to-one correlation between lexical and syntactic categorization, is typically attested in rigid languages, including many Indo-European languages. The latter option involves a mismatch between lexical and syntactic categories. A well-known example of a language for which this has been claimed is Tongan (see also the previous section). According to Broschart (1997), lexical categories in Tongan are defined in terms of morphological features, which have no relevance for syntactic distribution. Himmelman (2007) makes a very similar claim for Tagalog: According to him, there are a number of morphologically defined lexical categories in Tagalog, which are not projected into the syntax. There is some kind of correlation between these so-called *morpho-lexical categories* and purely ontologically defined categories, but this correlation is far from perfect. This means that the lexical categories of Tagalog are definitely *grammatical* categories, even though they are not *syntactic* categories.

Don and Van Lier (forthcoming) relate the issue of compositional and non-compositional semantic interpretation in flexible languages to a mismatch between lexical and syntactic categorization in these languages. They assume that, at a sufficiently abstract level of representation, all languages have a set of un-categorized roots with certain ontological meanings. In rigid languages, the operation that assigns these roots to a lexical category also automatically determines their phrase-structural destination: Once it is categorized, a lexical item can be used in only one syntactic function (without further measures). Flexible languages, in contrast, are characterized by the fact that they can lexically categorize and re-categorize their roots without affecting the phrase-structural possibilities of the output forms. Such non-syntactic (re-)categorization may or may not involve overt formal marking. The crucial point is that, since this type of lexical categorization occurs *prior* to syntactic categorization, it may involve non-compositional semantic interpretations. In contrast, syntactic categorization, i.e. the

ultimate insertion of lexical material into a syntactic function, should yield fully compositional semantic interpretations. Typically, the latter type of categorization involves syntactic frames made up of tense-aspect-mood (TAM) operators (for verbal/predicative usage), or determiners and/or case markers (for nominal/referential usage).

Consider a language like Samoan. In this language, according to Mosel and Hovdhaugen (1992: 77), “*the categorisation of words into nouns and verbs is not given a priori in the lexicon. It is only their actual occurrence in a particular syntactic environment which gives them the status of a verb or a noun*”. When a lexical item is preceded by a TAM particle it functions syntactically as a verb, and when it combines with one of the following particles it functions syntactically as a noun: *le* (determiner), *e* (ergative; absolutive is zero-marked), *o* (possessive), or *i* (marker of non-core arguments and locative-directional adjuncts).

The semantic interpretation of lexemes in verbal versus nominal syntactic function is often completely predictable, as is illustrated in examples (35) and (36). In example (35b), the action-denoting root *alu* ‘go’ occurs in nominal function and means ‘(the action of) going’. In example (36a) the object-denoting root *uō* ‘friend’ occurs in verbal function and is interpreted compositionally as ‘be friends’.

Samoan (Mosel & Hovdhaugen 1992: 73, 77)

(35) a. *E alu le pasi i Apia*
 GENR go DET bus DIR Apia
 ‘The bus goes to Apia.’

b. *le alu o le pasi i Apia*
 DET go POSS DET bus DIR Apia
 ‘the going of the bus to Apia.’

(36) a. *E uō Tanielu ma Ionatana*
 GENR friend Daniel and Jonathan
 ‘Daniel and Jonathan are friends.’

b. *E alofa Taniel i l=a=na uō*
 GENR love Daniel DIR DET=POSS=3SG friend
 ‘Daniel loves his friend.’

However, there are also many other possible semantic shifts, with less predictable interpretations. Some examples are given in (37). Note that the semantic relations between the paired forms make reference to actions and participants involved in those actions (actors, instruments, undergoers) or specific instances of those actions.

Samoan (Mosel & Hovdhaugen 1992: 82, 83)

- (37) a. *fana*: 'gun' (N) or 'shoot' (V)
 b. *lama*: 'torch' (N) or 'fish by torch light' (V)
 c. *gaoi*: 'steal' (V) or 'thief' (N)
 d. *solo*: 'move forward' (V) or 'procession' (N)
 e. *tusi*: 'write' (V) or 'letter/book' (N)

The semantic shifts that go with the conversions in the examples in (37) are not marked by any overt formal change. However, overt lexical derivation does occur in Samoan, in particular with the suffix *-ga*. Some examples are given in (38). Interestingly, the output forms of this process have irregular semantic patterns, which at least partly resemble the ones in (37). The derived forms may denote a participant, a location, or an institution associated with the action denoted by the base form, but also a specific instance of that action.

Samoan (Mosel & Hovdhaugen 1992: 195)

- (38) a. *amo* 'carry' - *āmo-ga* 'person(s) carrying loads'
 b. *a'o* 'teach' - *ā'o-ga* 'school'
 c. *tīpi* 'cut' - *tīpi-ga* 'surgical operation'
 d. *pule* 'control' - *pulē-ga* 'unit of church administration'

Since the semantic meanings of the derived forms in (38) are typically 'nominal', while those of the un-derived ones are 'verbal', *-ga* has traditionally been termed a nominalization suffix (Mosel & Hovdhaugen 1992: 194). However, according to Mosel (2004: 267) "*words carrying the so-called nominalization suffix can occur in the head position of a verb complex.*" In other words, *-ga* is *not* a syntactically categorizing derivational suffix: it creates a new lexical item with a particular semantic meaning, but does not determine the phrase-structural destiny of the item. This is illustrated in (39) with the form *to-gā* 'plantation', derived from the root *to* 'plant'. The derived form has an unpredictable semantic interpretation, which may be described roughly as:

‘location associated with the action denoted by the base form’. After entering into a compound with *niu* ‘coconut’, the derived form *togā* is inserted into a verbal syntactic function by combining it with the perfective particle ‘*ua*’. This syntactic categorization, as expected, yields a fully compositional semantic interpretation, namely ‘was a (coconut) plantation’.

Samoan (Mosel 2004: 267):

- (39) *‘Ua to-gā-niu ātoa le mea mau’u’epu’e*
 TAM plant-NMLZ-coconut whole ART place hill
 ‘The whole hilly place was now a coconut plantation.’

A very similar process occurs in Tongan, which has the suffix *-Canga* that can be applied to action-denoting, property-denoting, and object-denoting roots. According to Broschart (1997: 146), this suffix, just like Samoan *-ga*, is not a nominalizer in the phrase-structural sense, but rather a derivational marker whose output forms are flexible and mean something like “domain where something is or takes place”. However, the exact semantic interpretation of *Canga*-derived forms is far from fully predictable, as the examples in (40a-c) make clear:

Tongan (Broschart 1997: 146)

- (40) a. *pule* ‘govern’ – *pule-’anga* ‘government’
 b. *motu’a* ‘old’ – *motu’a-’anga* ‘reason for having aged’
 c. *api* ‘home’ – *api-tanga* ‘homestead’

Despite their ‘nominal’ semantics, Tongan *-Canga* derivations can be used as verbs by combining them with a TAM particle, as is illustrated in example (41):

Tongan (Broschart 1997: 145)

- (41) *’oku ikai ke pule’anga*
 PRS NEG SUBJ government
 ‘It does not belong to the government.’
 (lit. ‘it is not that it government-s.’)

I argue that the semantic shifts in the Samoan lexemes listed in (37) above are due to a process of derivation that must be analyzed as the zero-marked counterpart of *-ga* derivation. The same analysis applies to Tongan *-Canga* derivation, as compared with Tongan zero-derivation of the type illustrated

in (34b) above. Let us take the Samoan form *gaoi* in (37c) as a concrete example. The proposal is that there are two homophonous but structurally different forms *gaoi*: (i) an un-derived form with the meaning ‘steal’, and (ii) a zero-derived form with the meaning ‘thief’. Alternatively, the items in (37) can be analyzed as polysemous, i.e. as involving two meanings belonging to a single form. The crucial point, which is captured by both analyses, is that the cases in (37) are *not* instances of flexibility. As such they stand in contrast to the examples in (35) and (36), which involve syntactic rather than lexical categorization, and accordingly behave semantically fully regularly.

Notably, it is predicted that both forms of the examples in (37), just like *-ga* derivations and their base forms, are flexible, i.e. are open to usage in verbal and nominal syntactic functions, and that these usages will yield compositional semantic interpretations. Taking *gaoi* as an example again, the un-derived form with the meaning ‘steal’ would be interpreted as ‘to steal’ in verbal function, and as ‘(the act of) stealing’ in nominal function. The zero-derived form with the meaning ‘thief’ would be interpreted as ‘to be a thief’ in verbal function, and as ‘a/the thief’ in nominal function. Unfortunately, the available data are insufficient to confirm this prediction.

Interestingly, the Samoan *-ga* suffix can also be used as a syntactic marker, rather than as a marker of lexical derivation. As would be expected, in the former function its output receives fully compositional semantic interpretations, namely ‘the act of X-ing’, where X is the action denoted by the base. In addition, the two different levels at which *-ga* may operate correlate with a phonological difference: Lexical derivation involves vowel-lengthening in the base form, while syntactic derivation leaves the phonological structure of the base unaffected. The phonological and semantic differences between lexical and syntactic *-ga* formations are illustrated in (42), using the same examples as in (38) above.

Samoan (Mosel & Hovdhaugen 1992: 195)

(42)

Base	Lexical derivation	Syntactic derivation
<i>amo</i> ‘carry’	<i>āmo-ga</i> ‘person(s) carrying loads’	<i>amo-ga</i> ‘carrying’
<i>a’o</i> ‘teach’	<i>ā’o-ga</i> ‘school’	<i>a’o-ga</i> ‘education’
<i>tīpi</i> ‘cut’	<i>tīpi-ga</i> ‘surgical operation’	<i>tīpi-ga</i> ‘cutting’
<i>pule</i> ‘control’	<i>pulē-ga</i> ‘unit of church administration’	<i>pule-ga</i> ‘controlling’

A similar phenomenon is attested in Maori. This language also has a very high degree of lexical flexibility, and it has a suffix *-tanga*, which can be used as a marker of lexical and syntactic derivation (Bauer 1993: 254). Example (43) nicely shows both types of derivations from the same base form *tangi* ('cry') within a single clause. The lexical process yields the idiosyncratic interpretation 'funeral', while the syntactic formation is compositionally interpreted as 'the crying (of the people)'.

Maori (Bauer 1993: 48)

- (43) *I rongo au i te [tangi-hanga hotuhotu-tanga*
 T/A hear 1SG DIR.OBJ ART cry-NMLZ sob-NMLZ
O ngaa taangata i te tangi-hanga ki a
 POSS ART people at ART funeral LOC ART
Maui Poomare]
 Maui Poomare
 'I heard the people's crying at Maui Poomare's funeral.'

So far, we have seen that syntactic categorization of putatively flexible lexemes in languages like Samoan and Tongan involves fully compositional semantic interpretation. This is in accordance with Evans and Osada's compositionality criterion for 'true' flexibility. However, the data discussed above also show that such flexibility does not exclude the presence of lexical (re-)categorization processes. When such lexical processes are not overtly marked, they may be confused with flexibility. However, on the basis of the irregular semantic characteristics of these lexical processes, I have argued that they do not involve flexibility, but should rather be analyzed as instances of conversion.

We may now compare the Tongan and Samoan data discussed above with similar patterns found in other supposedly flexible languages. Consider for instance Tagalog, a language in which, according to Himmelmann (2005: 361), "*content words do not have to be sub-classified with regard to syntactic categories. They all have the same syntactic distribution, i.e. they all may occur as predicates [and] as (semantic) heads of noun phrases [...]*". Himmelmann analyzes voice-marking in Tagalog as a derivational process, which produces "*actor- or undergoer-oriented action expressions*" that are still flexible in terms of their syntactic possibilities. This process is characterized by "*pervasive formal and semantic idiosyncrasies*" (Himmelmann 2007: 288). Some examples of idiosyncratic semantic interpretations of voice-marked forms are provided in (44):

Tagalog (Foley 1998)

(44)	Base		Voice-marked form
a.	<i>abogado</i> ‘lawyer’	–	<i>mag-abogado</i> ‘become a lawyer’
b.	<i>bus</i> ‘bus’	–	<i>mag-bus</i> ‘ride a bus’
c.	<i>kamay</i> ‘hand’	–	<i>mag-may-an</i> ‘shake hands’
d.	<i>langgam</i> ‘ant’	–	<i>langgam-in</i> ‘be infested with ants’
e.	<i>lubid</i> ‘rope’	–	<i>lubir-in</i> ‘be made into rope’

Derivation with voice markers may also be irregular in terms of phonology: It may involve unpredictable deletion of root vowels, as illustrated in (45a), or sporadic insertion of /n/, as in (45b). There are also completely irregular forms, such as in (45c):

Tagalog (Himmelman3007: 288)

(45)	Base		Voice-marked form
a.	<i>kain</i> ‘consumption of food’	–	<i>kan-in</i> ‘to eat something’
b.	<i>tawa</i> ‘laugh, laughter, laughing’	–	<i>tawa-n-an</i> ‘to laugh at someone’
c.	<i>kuba</i> ‘getting, a helping’	–	<i>kun-in</i> ‘to get something’

As already mentioned, even though voice-marked forms denote actions, they can still be used in nominal syntactic function without further measures. As predicted, such syntactic categorization yields fully predictable semantic interpretations: An actor-voice form in nominal function denotes ‘the agent of the action’, a patient voice form in nominal function denotes ‘the patient of the action’, and so on (Himmelman 2007)²⁹. An example of a locative voice form in nominal syntactic function is given in (46):

²⁹ These Tagalog data show that a compositional interpretation of flexible lexemes is not necessarily ontologically equal across languages: The interaction with the specific feature of voice-marking in Tagalog accounts for the fact that in this language action-denoting forms are interpreted as objects or individuals in referential function, while in Samoan they are interpreted as denoting the action as a whole. See also the examples from Kharia further on.

is illustrated in example (48) below. In contrast, *lakad*₂ would mean ‘a/the walk’ when used as a noun, and ‘be a walk’ when used as a verb (as is in fact the case in example (47) above).

Tagalog (Himmelmann 2007: 281)

- (48) *Ma-busay* *ang* *lakad* *ng* *mákiná*
 STAT-orderliness SPEC walk GEN machine
 ‘The walking of the machine is good.’

These data show that, like the other languages discussed in this section, Tagalog displays ‘true flexibility’ in the sense that it freely uses simple and derived lexical material in all syntactic functions with compositional semantic interpretations. This flexibility is combined with either overtly or zero-marked lexical processes, which yield semantically non-compositional outcomes.

Finally, consider Kharia, a South Munda language that is unrelated to the Malayo-Polynesian languages discussed so far. Peterson (2005, 2006, forthcoming) claims that Kharia does not possess the lexical classes ‘noun’ and ‘verb’, but instead has two syntactically defined categories, which he calls ‘predicate’ and ‘complement’. According to Peterson (2006: 60), “*lexemes in Kharia do not appear to be either inherently predicates or their complement but can generally appear in both functions, without any overt derivational morphology.*” The categories ‘predicate’ and ‘complement’ are expressed through the combination of a so-called ‘content head’ with a functional head. Content heads may consist of any single lexical root, but also of multiple roots or complex phrases. A predicative functional head is expressed through enclitic markers for tense/voice and person, while a complement functional head is expressed through enclitic case marking (other than possessive/genitive).

The semantic interpretation of a content head in an actual utterance requires a combination of the intrinsic meaning of the root(s) that it consists of, the syntactic function in which it appears, and the morpho-syntactic distinctions belonging to that function (such as tense/voice in predicative function). According to Peterson (2006: 70), this combination yields “*entirely predictable*” semantic outcomes. In particular, if an object- or individual-denoting root is used as a predicate and marked for middle voice, it gets the meaning ‘to become X’ (where X is the denotation of the root). When the same type of lexeme in the same function is marked for active voice it means ‘to turn something into X’. Action-denoting roots that are used in

referential function get the meaning ‘(the act of) X-ing’ (where X is again the denotation of the root). Examples are given in (49) and (50):

Kharia (Peterson 2006: 60, 68)

- (49) a. *lebu dɛl=ki* b. *bhagwan lebu=ki*
 man come=M.PST God man=M.PST
 ‘The/a man came.’ ‘God became a man.’ (= Jesus)

- (50) a. *u kayom onɔor=kon raʔa=ya? ayo=qom,*
 this talk hear=SEQ Rata=GEN mother=3.POSS
 darhi=ya? saw-ʔay=qom gam=te:...
 Darhi=GEN spouse-woman=3.POSS say=ACT.PRS
 ‘Hearing this talking, Rata’s mother, Darhi’s wife says: ...’

- b. *ni lebu=ki khorɪ=ki=te kayom=ta=ki*
 story people=PL village.section=PL=OBL talk=M.PRS=PL
 ‘The people tell [this] story in the villages.’

The syntactic nature of *Kharia* categorization and the phrasal clitic-hood of the functional heads becomes even more evident when considering an example like (51), in which the complex content head *bidesa? ebukiya? rupraŋ* ‘the appearance of people from abroad’, is used in the syntactic function of predicate. This involves combining the content head with a functional head, which is expressed by clitic forms for voice, tense and person, and attaches to the phrase’s final element.

Kharia (Peterson 2006: 85)

- (51) *bharat=ya? lebu=ki [bides=a? lebu=ki=ya?*
 India=GEN person=PL abroad=GEN person=PL=GEN
 rupraŋ]=ki=may.
 appearance=M.PST=3PL
 ‘The Indians took on the appearance of people from abroad.’

Just like the other flexible languages discussed above, *Kharia* also has lexical processes that involve semantically and phonologically unpredictable outcomes. Consider the derivational process marked by *nasal-vowel-infixation* in *Kharia*. The output forms of this process may denote an object, an instrument, or a location typically involved in the action denoted by the

base form, as well as a specific instance or result of that action. With respect to phonology, the infix vowel has the same quality as the vowel preceding the nasal, but the infix nasal is of indeterminate quality: it is usually realized as /n/, but /m/ is also occasionally found (Peterson 2006: 79). Some examples of the process are given in (52):

Kharia (Peterson 2006: 80)

(52)	Base		Lexical derivation	
a.	<i>bel</i>	‘spread out’	<i>be-ne-l</i>	‘bedding’
b.	<i>bui</i>	‘keep, raise’	<i>bu-nu-i</i>	‘pig’
c.	<i>jo?</i>	‘sweep’	<i>jo-no-?</i>	‘broom’
d.	<i>jin</i>	‘touch’	<i>ji-ni-b</i>	‘touch’
e.	<i>rab</i>	‘bury’	<i>ra-na-b</i>	‘burial ground’
f.	<i>kuj</i>	‘dance’	<i>ku-nu-j</i>	‘dance’
g.	<i>mu?siŋ</i>	‘rise’	<i>mu-nu-?siŋ</i>	‘east’

According to Peterson, and in line with the lexical (zero-)derivational processes discussed for other flexible languages, *Kharia* infix forms can be used freely in nominal and verbal syntactic function, despite their ‘nominal’ meanings. As is the case for any other type of content head, this yields entirely compositional semantic interpretations. Thus, a derived form like *bunui* ‘pig’ (see (52b) above) will get the meaning ‘to become a pig’, when used as a predicate in middle voice, but ‘to turn something//someone into a pig’ when used as a predicate with active voice. In nominal syntactic function it will mean ‘a/the pig’. Obviously, the predicative uses of a form like *bunui* are pragmatically rather odd, and will therefore be seldom attested, if at all. However, examples like those in (49b) above make clear that, given the appropriate pragmatic context, the use of object- or individual-denoting lexemes in verbal syntactic function is grammatically totally unproblematic.

To summarize, I argued in this section that the patterns of semantic interpretation of lexical material in presumably flexible languages can be understood when taking into account the mismatch between lexical and syntactic categorization in these languages. In flexible languages, lexical categorization is not relevant for phrase structure. This stands in contrast to typical rigid languages, in which there is a one-to-one correlation between lexical categorization and syntactic categorization. Possibly, these two scenarios represent the extreme ends of a scale or cline. This would mean that there are also languages in which syntactic categorization takes place

somewhere in between the level of simple lexical root and the level of phrase structure. In Chapter 8 I will discuss this idea and its repercussions for the present study in more detail.

2.5.3 The criterion of exhaustiveness: How flexible should a flexible language be?

The second criterion proposed by Evans and Osada (2005) concerns the amount or the proportion of flexible lexical items that is required in order for a language to qualify as flexible. This criterion is described as follows:

“It is not sufficient to find a few choice examples which suggest word class flexibility. Since word classes are partitionings of the entire lexicon, equivalent statements need to hold for all relevant words in the lexicon that are claimed to have the same class”. (Evans & Osada 2005: 378)

It seems, in fact, that there is considerable agreement between scholars on this point. For instance, Hengeveld et al. (2004: 538) remark that “*in order for a language to qualify as flexible, it has to show systematic flexibility (...) for an entire class of lexemes.*” And Baker (2003: 117) makes a similar point saying that “*an important typological difference exists only if categorial ambiguity extends to an entire open class of inflectionally similar words*”.

However, Evans and Osada (2005: 378) rightly observe that “*in practice (...) it is difficult to check every one of tens and thousands of lexical items*”. In spite of this rather obvious practical restriction, it seems that in many reference grammars and other descriptive sources quite unambiguous statements can be found about the pervasiveness of multi-functionality of (certain groups of) lexemes, be it not exactly in the quantified fashion that Evans and Osada aim at. This is clearly illustrated by the quotations in the previous subsection about such languages as Samoan, Tagalog, and Kharia. In Chapter 5 I will further discuss the criterion of exhaustiveness and its application to the relevant languages of the sample.

2.5.4 The criterion of equivalent combinatorics: Bi-directional flexibility

The third and final criterion for flexibility proposed by Evans and Osada is the criterion of *equivalent combinatorics*. It states that

“[M]embers of what are claimed to be merged classes [i.e. flexible classes, EvL] should have identical distributions [...]. All members of both

putative classes should be equally acceptable in both primary functions as argument and predicate”. (Evans & Osada 2005: 366-367)

A corollary of this criterion is that flexibility should be bi-directional:

“To establish that there is just a single [i.e. flexible, EvL] word class, it is not enough for Xs to be usable as Ys without modification: It must also be the case that Ys are usable as Xs.” (Evans & Osada 2005: 375)

The part of this criterion that makes reference to the actual possibility to use a flexible lexeme in more than one function seems quite unproblematic. Probably anyone would agree that this is an essential part of any definition of flexibility. Rather less straightforward is the part of the bi-directionality criterion that refers to equal acceptability of every lexeme in all functions. In particular, the frequency with which flexible lexemes appear in predicative or referential function is bound to be influenced by semantic and pragmatic factors. Regarding this point, I follow Hengeveld and Rijkhoff, who argue as follows:

“There is no reason to assume that all our concepts are equally symmetrical with respect to predicating and referring functions in a particular language. We may expect certain flexible lexemes to occur more as predicates than as arguments, whereas other lexemes are used more often as arguments than as predicates.”
(Hengeveld & Rijkhoff 2005: 412)

In other words, the semantic denotation of a root affects its relative markedness with respect to a particular propositional function. This is in fact convincingly shown by Croft’s prototype approach to PoS classes, as discussed in section 2.4.2. Also, recall in this respect the examples from Samoan, Tongan, and Kharia mentioned in the previous subsection, which involve semantically ‘nominal’ but syntactically flexible (derived) lexemes. Crucially, the fact that the use of a lexeme with a particular semantic meaning in a particular function is pragmatically marked and therefore infrequently attested does not mean that there is a syntactic restriction that excludes such usage. In Chapter 5 I will further illustrate this point for the relevant languages in the sample.

A final aspect of bi-directionality is specifically related to noun-adjective flexibility. There are languages in which property-denoting lexemes

(i.e. semantic adjectives) can be the sole element in a referential phrase³¹.

However, this does not necessarily warrant the conclusion that there is no lexical distinction between adjectives and nouns. Often, such cases of apparent noun-adjective flexibility in fact involve so-called *absolutely used adjectives*, i.e. adjectives which function as the modifier of a head that is understood from the discourse context and therefore remains unexpressed.

Moravcsik (2001) distinguishes two ways in which property-denoting lexemes can function as the sole element in a referential phrase, namely referentially *dependent* and referentially *independent*. In the former case, the head noun must be available from the discourse or situational context. In the latter case, the intended referent is invariant, such as ‘people’ in (53) below. This possibility of referentially *independent* use of property-denoting words seems to be a requirement for ‘true’ noun-adjective flexibility.

Hungarian (Moravcsik 2001: 339)

- (53) *Sok gazdag van itt*
many rich is here
‘There are many rich (people) here.’

However, it is often problematic to determine whether referentially independent usage of property-denoting lexemes as the head of a referential phrase is a real option in an individual language. This is especially so when the analysis is based on isolated examples from written sources. Sometimes however, evidence in favour of an independent usage analysis can be found in the domain of agreement. In particular, the understood head may trigger the expression of agreement morphology on the absolutely used adjective. This happens for instance in Spanish, as shown in example (54), where *moderna* takes a feminine singular agreement suffix because it modifies the un-expressed but understood head *la casa* ‘the house’.

Spanish (Hengeveld 1992: 61-62)

- (54) *prefiero es-a moderna-a*
prefer:1s.PRS that-FEM.SG modern-FEM.SG
‘I prefer that modern (one).’ [e.g. the house]

³¹ Schachter & Shopen (2007: 17) use the term ‘adjectival-noun languages’ for this type of language.

Note furthermore that, if one takes seriously the criterion of bi-directionality, this implies that, in order to establish ‘true’ noun-adjective flexibility, it should be possible not only to use property-denoting lexemes as heads of referential phrases, but also to use object- or individual-denoting lexemes (i.e. semantic ‘nouns’) without further measures as modifiers in referential phrases. Again, this is not always easily determinable, since the distinction between nominal apposition or composition on the one hand, and attributive modification on the other hand may be unclear. In Chapter 5, where the PoS systems of the individual sample languages are categorized, I will discuss the relevant cases in more detail.

2.5.5 Lexical flexibility: A summary

In this section, I discussed the three criteria for lexical flexibility proposed by Evans and Osada (2005): Semantic compositionality, exhaustiveness, and equivalent combinatorics. Especially the first criterion has given rise to disagreement in the literature: While Evans and Osada contend that ‘true’ flexibility involves fully compositional semantic interpretation of lexemes in particular functional environments, other scholars argue that the occurrence of semantic idiosyncrasies is not an argument against flexibility. I have shown that a number of languages, which figure prominently in discussions on lexical flexibility, display both semantically compositional and non-compositional (zero-marked) derivational processes. The difference between these two types of processes lies in the fact that one pertains to syntactic, and the other to lexical (re-)categorization. The special characteristic of flexible languages, then, seems to be the lack of a one-to-one correspondence between these two levels of categorization.

As regards the remaining two criteria, the one concerning exhaustiveness seems theoretically unproblematic, but is hard to verify empirically. Finally, with respect to the criterion of equivalent combinatorics, it was argued that while bi-directionality is probably a central notion in flexibility, this does not mean that flexible lexemes are expected to appear equally frequently in all of their possible functions.

2.6 Summary

This chapter presented a number of functional-typological approaches to lexical categorization. I started out with a brief outline of some early studies, and the criteria that were used in them to define PoS classes. Subsequently, I provided a detailed discussion of Hengeveld’s PoS theory. In its most

recent version, this theory proposes an implicational map consisting of four propositional functions defined by two binary parameters: *predication-reference*, and *head-modifier*. Hengeveld's definitions of PoS classes are based on the presence versus absence of structural coding accompanying the use of lexemes in each of these four possible functions. The dominance relations holding between the two parameters and between their respective values predict a number of PoS systems that are expected to occur in actual languages.

Hengeveld's theory excludes two types of defining criteria for PoS, namely morphological criteria (or behavioral potential) and semantic criteria. The advantage of this approach is that it provides the possibility to compare PoS classes and PoS systems across languages. However, exactly this aspect of the Hengeveldian method has been criticized, most elaborately by Croft (2000, 2001, 2005). The latter objects that Hengeveld's theory, since it ignores morphological and semantic data, amounts to methodological opportunism and misses out on cross-linguistic universals to be found in these two areas. On the other hand, the consequence of Croft's method, namely taking into account the full set of distributional data, is that lexical categories, like any other type of language-specific formal category, can no longer be compared across languages. Croft's solution to this problem is to define PoS as typological prototypes, i.e. as unmarked combinations of semantic classes and pragmatic functions, which correlate cross-linguistically with relative degrees of formal marking.

Another major point of discussion concerns the status of so-called flexible languages, and especially the role of semantics in defining what lexical flexibility really entails. Evans and Osada (2005, and cf. Croft 2005) propose that in truly flexible languages the interpretation of lexical material in different functional environments should be fully compositional. I discussed data from supposedly flexible languages, which show that this criterion is indeed met. At the same time, however, such languages display lexical processes of (zero-)derivation and polysemy, and these are semantically (as well as phonologically) non-compositional. I argued that the presence of both compositional and non-compositional semantic (re-)categorization in flexible languages can be understood in terms of the fact that these languages, unlike in the more familiar rigid languages, lack a one-to-one correlation between lexical and syntactic categorization.

In the next chapter, I will use Hengeveld's method for lexical classification to develop a typology for a different set of constructions, namely dependent

clauses. Therefore, it seems appropriate at this point to emphasize that I fully acknowledge the limited scope of the Hengeveldian approach in terms of the type and amount of analytic primitives that it takes into account. However, unlike Croft, I do not regard this as methodological opportunism. Depending on the aim of one's research, it seems perfectly defensible to focus on certain linguistic facts while ignoring others, as long as this is made explicit. As already mentioned, the lexical categories defined within Hengeveld's framework are claimed to be cross-linguistically comparable only as far as his very specific set of criteria goes. And, as Cysouw (2007: 227) puts it: "*Cross-linguistic identity is always just a matter of granularity of analysis*".