On flexible and rigid nouns

Jan Rijkhoff

Department of Linguistics, University of Aarhus, Denmark

This article argues that in addition to the major flexible lexical categories in Hengeveld's classification of parts of speech systems (Contentive, Non-Verb, Modifier), there are also flexible word classes within the rigid lexical category Noun (Set Noun, Sort Noun, General Noun). Members of flexible word classes are characterized by their vague semantics, which in the case of nouns means that values for the semantic features Shape and Homogeneity are either left un-determined or they are specified in such a way that they do not quite match the properties of the kind of entity denoted by the flexible item in the external world. I will then argue that flexible word classes constitute a proper category (i.e. they are not the result of a merger of some rigid word classes) in that members of flexible word categories display the same properties regarding category member-ship as members of rigid word classes. Finally this article wants to claim that the distinction between rigid and flexible noun categories (a) adds a new dimension to current classifications of parts of speech systems, (b) correlates with certain grammatical phenomena (e.g. so-called number discord), and (c) helps to ex-plain the parts of speech hierarchy.

1. Introduction

The central claim of this paper is that nouns that are used cross-linguistically to refer to a single, concrete object can be divided into flexible and rigid subtypes. A rigid noun has a narrow content, which means that the values for certain semantic features (Shape, Homogeneity) match the properties of a concrete object in the physical world. A flexible noun, on the other hand, is semantically vague: values for these features are either left unspecified or they are specified in such a way that they do not quite cor-respond with the properties of a single, concrete object in the external world. Thus, in addition to major flexible word categories such as Contentive or Non-verb (Hengeveld 1992; Hengeveld et al. 2004; Hengeveld and Rijkhoff 2005), we also find flexible nomi-nal (sub)categories within the category Noun (Section 4).

It is then claimed that a flexible word class is not a merger of some rigid word classes, but constitutes a proper category in its own right in that category membership

can be described in terms of subsective and intersective gradience (Section 5). This article generally demonstrates the importance of the rigid/flexible distinction in classifications of parts of speech systems. Apart from the fact that the distinction between rigid and flexible noun categories refines current classifications of parts of speech systems, it is shown that flexible categories correlate with certain grammatical phenomena (such as so-called number discord) and that the features that are used to define flexible and rigid noun categories also play an important role in the parts of speech hierarchy (Verb > Noun > Adjective > Adverb; Section 6).

2.   Parts of Speech systems

Whereas words in many languages can be categorized in terms of the traditional word classes (Verb, Noun etc.), there are also quite a few languages which are deemed to have a major lexical word class whose members do not seem to belong to any of the traditional (rigid) word classes like Verb, Noun or Adjective. This is captured in Hengeveld's classification of parts of speech (PoS) systems, which has a major division between languages with a flexible PoS system and languages with a rigid PoS system (Figure 1). A simplified version of this classification is given in Figure 1. In languages with a flexible PoS system, some or all of the functions of a lexeme in an actual linguistic expression (i.e. verbal, nominal, adjectival, adverbial function) are performed by the same group of lexemes (Types 1–3). In languages with a rigid PoS system (Types 4–7) these functions are distributed over distinct, non-overlapping groups of lexemes. One could also say that a flexible lexeme (a contentive, a non-verb or a modifier) is semantically vague in the sense of Cruse (1986: 51) in that it has a general meaning, which covers two or more of the specific possibilities.

Figure 2 shows how the seven PoS systems in Hengeveld's classification are distributed in a representative sample of the world's languages (n=52). Since I will largely ignore adverbs, the number of PoS systems considered here can be reduced to five. Notice that I have included intermediate types 3/4 and 4/5 which have a minor class of adjectives and nouns, respectively (see also Hengeveld et al. 2004 and Smit 2001 for a more detailed presentation of languages with an intermediate PoS system).

FLEXIBLE PoS	*1*	*Contentive*				
	2	Verb			*Non-verb*	
	3	Verb		Noun		*Modifier*
Rigid PoS	*4*	Verb	Noun		Adjective	(manner) Adverb
	5	Verb		Noun		Adjective
	6	Verb			Noun	
	7	Verb				

Figure 1. Hengeveld's classification of parts of speech systems (slightly modified).

Type 1: *flexible*	'*Contentive*'	**Samoan**
Type 2: *flexible*	V – '*Non-verb*'	**†Hurrian, Imbabura Quechua, Turkish**
Types 3:	V – N – A	Abkhaz, Alamblak, Basque, Berbice Dutch Creole, Bukiyip (=Mountain Arapesh), Burush-aski, Dutch, Georgian, Guaraní, †Hittite, Hmong Njua, Hungarian, Ika, Kayardild, Ket, Nama Hottentot, Nasioi, Ngalakan, Ngiti, †Sumerian, Wambon
Type 3/4	V – N(– A)	Babungo, Bambara, Chukchi, Gude, Kisi, Oromo, Pipil, Sarcee, Tamil
Type 4	V – N	Burmese, Galela, Hixkaryana, Koasati, Korean, Krongo, Lango, Mandarin Chinese, Nivkh (=Gilyak), Nung, Nunggubuyu, Tsou, Vietnam-ese, West Greenlandic
Type 4/5	V(– N)	Cayuga
Type 5	V	
Type unknown		*†Etruscan, †Meroitic, Nahali*

Figure 2. The distribution of Parts of Speech systems [without manner adverbs] in a representative sample of the world's languages (Rijkhoff 2004: 18).

Samoan is the only language in the sample with an extremely flexible PoS system (see Hengeveld & Rijkhoff 2005 and Broschart & Dawuda 2000 on other languages without a rigid verb/noun distinction). It has a single word class, whose members can be used as

- head of the clause (verbal function), as in (1a),
- head of the 'noun phrase' (nominal function; notice that 'noun phrase' is actually a misnomer here), as in (1b), or
- modifier of these heads (adverbial or adjectival function).

> **Type 1**:
> Samoan (Austronesian; Mosel and Hovdhaugen 1992: 80, 73, 74)

(1) a. 'Ua lā   le   aso   b. 'Ua   mālosi le   lā
        PERF sun ART day   PERF strong ART sun
        'The sun is shining today.'   'The sun is strong.'
        (lit. 'The day suns')   (lit. 'The sun strongs.')

Galela and Hausa are both examples of languages with two rigid classes: verbs and nouns. To express adjectival notions Galela uses participial forms of stative verbs, whereas Hausa employs an attributive NP headed by an abstract noun.

Type 4:
Galela (Papuan; van Baarda 1908: 35)

(2) awi ḍohu **i lalamo**
     his foot   it be_big:PRT
     'his big foot'

Hausa (Afro-Asiatic; Schachter 1985: 15)

(3) mutum **mai alheri**   / **arzaki**   / **hankali**
     person with kindness / prosperity / intelligence
     'a kind/prosperous/intelligent person'

Experts disagree whether there are languages with a single lexical word class that con-
sists of just verbs (cf. Hengeveld and Rijkhoff 2005: 408), but Cayuga and Tuscarora
seem to come rather close to a Type 5 language:

Type 4/5:
Cayuga (Iroquoian; Sasse 1993: 657)

(4) a-hó-htọ:'   ho-tkwę't-a'   nę:kyé họ-kweh
     PAST-it:to_him-become_lost it:him-wallet-be this   he:it-man
     'This man lost his wallet.'
     (the literal meaning, however, would be something like (Sasse: ibid.): "it
     became lost to him, it is his wallet, he is this man" or "it LOSTed him, it
     WALLETs him, the one who MANs")

Tuscarora ((Iroqiouan: Mithun Williams 1976: 32)

(5) rakwá:tihs   wahratkáhthoʔ   katéskrahs
     ra-kwatihs   wa-hr-at-kahtho-ʔ   ka-teskr-ahs
     M-young   AOR-M-look_at-PNCT NONH-stink-SER
     he_is_young he_looked_at_it   it_stinks
     'The boy looked at the goat'

It appears that flexibility in the PoS system correlates with a variety of grammatical
phenomena, such as: fixed word order, suppletion, absence of gender and number
(transnumerality), no copula support with bare predicates, and absence of transitive
lexemes (Rijkhoff 2003, 2004; Hengeveld et al. 2004; Hengeveld and Rijkhoff 2005).

3.   The vague semantics of flexible lexemes

Hengeveld's classification of PoS systems (Figure 1) contains three major word classes
whose members are semantically vague: *Contentive*, *Non-verb* and *Modifier*. In the
sample, Samoan is an example of a language with contentives. Apparently "there are
no lexical or grammatical constraints on why a particular word cannot be used in the
one or the other function" (Mosel and Hovdhaugen 1992: 73–4, 77):

> "… the categorization of full words is not given a priori in the lexicon. […] It is only their actual occurrence in a particular environment which gives them the status of a verb or a noun. […] What is given in the lexicon, is not a particular word class assignment, but the potential to be used in certain syntactic environments as a noun or a verb.[1]
> […] all full words which function as noun and verb phrase nucleus can also be used as attributive modifiers."

In other words, it is the use of a vague lexeme in a certain context (the actual linguistic expression) that brings out certain parts of its meaning, giving the category neutral item a verbal, nominal, etc. flavour (see below on *coercion*). Inspired by Wilkins's account of noun semantics in Arrernte (Wilkins 2000), an outline of what may happen when a member of a flexible word class is used in an actual linguistic expression was presented in Hengeveld et al. (2004) and Hengeveld and Rijkhoff (2005). In this approach the vague meaning of a flexible lexeme consists of meaning components A B C D E etc. (where A B C etc. can be, for example, features, definitions, descriptions, knowledge structures). By placing the contentive in a particular syntactic slot the speaker *highlights* those meaning components that are relevant for a certain lexical (verbal, nominal, etc.) function, *downplaying* other meaning components. A very simplified representation of this process may look as follows:

(6)   Meaning components of Samoan *lā* (A B C D E)

	A	B	C	D	E	… Highlighted properties of *lā*:
Slot: head of clause	+		+		+	A C E => verbal meaning (*lā* 'be_sunny')
Slot: head of 'NP'		+		+	+	B D E => nominal meaning (*lā* 'sun')
Slot: modifier of 'noun'		+	+	+		B C D => adjectival meaning (*lā* 'sunny')

Notice that no meaning is added, i.e. there is no semantic increment. Rather, the context highlights certain meaning components that are already present in the meaning of the lexeme, giving the flexible item its verbal, nominal or adjectival flavour. It will be argued below that the same approach ('coercion') can be used to account for the meaning of a flexible noun in an actual utterance.

4.   Flexible and rigid nouns

The distinction between flexible and rigid PoS systems in *Figure* 1 is also relevant within the rigid word classes verb and noun. In some languages verbs are semantically specified for the familiar event-type (*Aktionsart* 'mode of action') features such as Telicity; in other languages verbs are vague (underdetermined or underspecified) with

FLEXIBLE PoS	1	Contentive			
	2	Verb		Non-verb	
Rigid PoS	3a	Verb	Noun^N ('narrow content')		Adjective
	3b	Verb	*Noun^W ('wide content')*		Adjective
	4a	Verb		Noun^N ('narrow content')	
	4b	Verb		*Noun^W ('wide content')*	
	5	Verb			

Figure 3. Parts of speech systems with rigid ('narrow content') and flexible ('wide content') nouns (cf. Figure 6).

respect to such features (Ikegami 1985; see also e.g. Bisang 1996: 520; Bohnemeyer 2002). Here, however, we will only be concerned with nouns (i.e. languages of type 3–4 in Figure 2). I will argue below that nouns that are used to refer to a single concrete object have a *narrow semantic content* ($Noun^N$) if the values for the semantic features Shape and Homogeneity match the properties of a single object in the physical world. This is not so in the case of nouns with a *wide semantic content* ($Noun^W$), which are semantically vague. In the case of a flexible noun we find that the values for certain features are either left unspecified (Homogeneity) or are specified in such a way that they do not quite match the properties of a concrete object in the real world (Shape). For example, a language such as Thai, which, unlike e.g. Dutch, uses transnumeral nouns for concrete objects such as 'table', 'car', or 'knife', can be said to have nouns with a wide semantic content (Hundius and Kölver 1983: 181f):

> "[…], Thai nouns do not in themselves contain any numerical or referential indi-
> cations. […] they are purely conceptual labels which, in order to be appropriately
> related to objects of the non-linguistic world, always and in principle stand in
> need of interpretation which has to be inferred from both linguistic and non-
> linguistic context."

Thus, in addition to languages with flexible word classes such as Contentive or Non-verb, whose members can be characterized as semantically vague, we also find that there are languages with flexible nouns that are characterized by vague semantics (Figure 3).

4.1 Noun classification: *Seinsarten*

A cross-linguistic classification of nouns requires that languages are studied against each other in terms of the same property (Song 2001: 11). So even if we have estab-lished that certain languages have a true class of nouns according to the criteria em-ployed in Hengeveld's classification of PoS systems (Hengeveld 1992; Hengeveld and Rijkhoff 2005), we still have to make sure that we make a valid comparison between nouns in different languages. To achieve this, I have assumed that there are concrete objects in the external physical world ('common sense metaphysics') and that they can be used as reference points (*tertium comparationis*). In other words, I have investigated

properties of nouns in a representative sample of the world's languages that are used to refer to a single concrete object in the external, physical world. Obviously this excludes languages with a PoS system that does not include the category Noun, i.e. languages of type 1, 2 and 5 in *Figure* 3. I have also ignored special or exceptional cases such as (1) derived nouns (e.g. *sainthood*), (ii) *pluralia tantum* (e.g. *scissors*, *measles*), or (iii) *singularia tantum* (e.g. *dust*). Singularia tantum are often actually mass nouns (*snow*), verbal nouns/infinitives (<u>Dutch</u> *het verkopen van …* lit. 'the selling of …'), abstract nouns (*thirst*, *hate*), proper names (*Max*, *Lisa*) and names for chemical elements (*helium*).

Since a concrete object is typically characterized by a definite spatial outline (i.e. it has a shape) and an internal structure containing parts or components (i.e. it is not a homogeneous entity), this investigation mainly focused on the two semantic features Shape and Homogeneity (likepartedness). As only entities with a definite outline can be counted, it was assumed that nouns that can be quantified directly have a positive value for the feature Shape. Conversely, nouns that require the presence of some kind of classifier (or 'individualizer'; Lyons 1977: 462) were deemed to have a negative value for the feature Shape (but see note 14). Thus, both Dutch and Oromo use +Shape nouns, even though in Oromo nouns are transnumeral whereas in Dutch they aren't (Andrzejewski 1960: 71):

> "The vast majority of [Oromo nouns] are associated with neither plurality nor singularity, i.e. the forms themselves give us no information as to whether what is denoted by them is one or more than one. When such forms are used, only the context can provide us with information about the number of what is denoted."

> <u>Oromo</u> (Stroomer 1987: 59): SET NOUN
(7) gaala   lamaani [noun+numeral; no plural]
camel(s) two
'two camels'

I have called nouns of the kind attested in Oromo 'set nouns' since they seem to denote a set, which can have any cardinality (including 'one').[2] Basically there are two kinds of sets: singleton sets with only one member and multiple sets with more than one member. By contrast, in Dutch the unmarked noun denotes a singular object and plural marking is obligatory with and without a numeral higher than 'one'.

> <u>Dutch</u>: SINGULAR OBJECT NOUN
(8) twee boek-en [numeral+noun+plural]
two   book-PL
'two books'

Thus, it seems that in Dutch the numeral specifies the number of individual objects, whereas in Oromo the numeral specifies the size of a set (for a more detailed comparison between singular object nouns and set nouns, see Rijkhoff 2003: 219–220; Rijkhoff 2004: Chapter 2).

Thai is one of the languages with nouns that cannot be quantified directly. According to Hundius and Kölver (1983: 166, 181–182) this is because Thai nouns: "[…] purely denote *concepts* and, for this reason, are incompatible with direct quantification".

Thai (Hundius and Kölver 1983: 172): SORT NOUN
(9)   rôm   săam khan [noun+numeral+sortal_classifier]
      umbrella(s) three CLF:long, handled object
      'three umbrellas'

The appearance of a classifier is required because the meaning of a sort noun does not seem to include the notion of spatial boundedness or discreteness. They are labeled sort nouns because they occur with so-called sortal or numeral classifiers (perhaps a better name would be 'concept nouns'). Sort nouns must not be confused with mass nouns like 'water' or 'sugar', which occur with a MEASURE WORD or MENSURAL CLASSIFIER ('a litre/bucket/bottle/cup(ful) of water'); on the difference between sortal and mensural classifiers see e.g. Hundius and Kölver 1983; Cheng and Sybesma 1999; Rijkhoff 2004: 48; Huang and Ahrens 2003).

Thai (Hundius and Kölver 1983: 170): MASS NOUN
(10) náamtaan săam thûaj [noun+numeral+mensural_classifier]
      sugar   three cup
      'three cups of sugar'

Yucatec also uses classifiers, but in this language there does not seem to be a fundamental distinction between sortal and mensural classifiers (Lucy 1992: 43, 74, 83):[3]

"Interpretatively, in Yucatec all nouns […] are neutral with respect to logical unit or shape."

"Outside of the restriction on compatibility with other classifiers, little in the grammar of Yucatec appears to hinge on, or correlate with, this 'sortal' […] versus 'mensural' distinction, and it is difficult to know what status it should be given."

Yucatec (Lucy 1992: 74): GENERAL NOUN
         a/one-CLF   banana(s) [numeral+general_classifier+noun]
(11) a. 'un-tz'íit   há'as   'one/a 1-dimensional banana (i.e. the fruit)'
      b. 'un-wáal   há'as   'one/a 2-dimensional banana (i.e. the leaf)'
      c. 'un-kúul   há'as   'one/a planted banana (i.e. the plant/tree)'
      d. 'un-kúuch há'as   'one/a load banana (i.e. the bunch)'
      e. 'um-p'íit   há'as   'one bit banana (i.e. a bit of the fruit)'

Homogeneity is best illustrated by masses (e.g. 'water') or collective entities (e.g. 'family'). What masses and collectives have in common is that they both consist of parts, called *portions* and *members* respectively, which can be referred to by same name. The portions of a mass such as 'water' can also be referred to as 'water' and, conversely,

when water is added to water the final product is also referred to as 'water' (i.e. masses are divisible and agglomerative entities). Similarly, any number of members of a collective entity such as 'family' higher than one are also 'family' and when a new member is added to the collective, we can still refer to it as 'family' (i.e. collectives are also divisible and agglomerative entities; see Kuhn 1982 for a detailed treatment of the notion 'collective').

Singular object nouns and sort nouns, on the other hand, only apply to the whole object. For example, the Dutch noun *fiets* 'bicycle' cannot be used to refer to any of its parts or components (e.g. the wheel, the chain, the handle bar). Set nouns and general nouns seem to be neutral with respect to the feature Homogeneity. A set noun can refer to a single concrete object or to a multiplicity of concrete objects. Similarly a general noun can be used to refer to, for example, a single concrete object or to a mass entity (see Yucatec example above). This classification of nouns is captured in Figure 4, in which each noun category specifies a different *Seinsart* ('mode of being'; comparable to *Aktionsart* 'mode of action' in verb semantics; see Rijkhoff 2004 (Chapter 2) for details).[4]

	−HOMOGENEITY	+HOMOGENEITY
−SHAPE	*General Noun*	
	Sort Noun	Mass Noun
+SHAPE	*Set Noun*	
	Singular Object Noun	Collective Noun

Figure 4. Noun categories: *Seinsarten* (Rijkhoff 2004: 54)

In sum, across languages four noun categories are used to refer to a single concrete object in the physical world (printed in italics in Figure 4):[5]

Singular object noun (+Shape, −Homogeneity): the noun designates a property that is characterized as having a definite outline and as being **non**-agglomerative;
Set noun (+Shape): the noun designates a spatial property that is only characterized as having a definite outline;
Sort noun (−Shape, −Homogeneity): the noun designates a spatial property that is characterized as **not** having a definite outline and as being **non**-agglomerative;
General noun (−Shape): the noun designates a spatial property that is only characterized as **not** having a definite outline.

Apparently no language uses mass nouns or collective nouns, i.e. nouns with a positive value for the feature Homogeneity, to refer to a single, concrete object:[6]

Mass noun (−Shape, +Homogeneity): the noun designates a spatial property that is characterized as **not** having a definite outline and as being agglomerative (parts: portions);

Collective noun (+Shape, +Homogeneity): the noun designates a spatial property that
is characterized as having a definite outline and as being agglomerative (parts:
members).

Table 1 shows the distribution of singular object nouns, set nouns, and sort nouns in
the languages listed in *Figure* 2. The fourth nominal category that is used to refer to
a single concrete object (General Noun) is attested in Yucatec, which is outside the
original sample.[7] Notice that a language may employ members from more than one
noun category to refer to a single concrete object (e.g. singular object nouns to refer
to human or animate entities and set nouns to refer to non-human or inanimate enti-
ties; see note 7). Notice furthermore that the sample contains a number of languages
that defy straightforward classification (for details see Rijkhoff 2004: 28–56). There are
basically four problems. The first problem relates to the fact that in some languages
number is not marked on the noun but on the noun phrase (as in Alamblak, Bambara,
Guaraní and Nama Hottentot; hence these languages have no value in *Table* 1). Then
there are languages in which the numeral is not a noun modifier, but a sentence adverb
(e.g. Hixkaryana), the head of a non-noun modifying clause (e.g. Koasati) or some
other kind of non-attributive expression (such as the body-part counting systems of
the Papuan languages, where counting involves touching the relevant body-part). Fi-
nally, nouns may be in transition so that they display properties of 'old' noun category
X and at the same time exhibit features of 'new' noun category Y (e.g. Nivkh; see also
note 5).

The information contained in Table 1 shows, among other things, that flexible
noun categories (involving nouns with a 'wide semantic content' such as set nouns or
sort nouns) are attested in more languages than the rigid category of Singular Object
Noun (whose members are characterized by a 'narrow semantic content') and that sort
nouns seem to be restricted to languages with two rigid word classes: verbs and nouns
(Type 4; the values of these Type 4 languages are highlighted by the box in *Table* 1). We
will return to the second point in Section 6.

4.2 The semantics of rigid and flexible nouns: 'Narrow' versus 'wide' semantic content

It was mentioned above that a singular object noun (which belongs to a rigid noun
category) can be said to have a narrow semantic content in that the values for the fea-
tures Shape and Homogeneity are set in such a way that they match the properties of
a concrete object in the physical world (+Shape, −Homogeneity). For example, a chair
has a spatial outline and consists of certain components. The values for the features
Shape and Homogeneity of the English noun 'chair' (+Shape, −Homogeneity) seem
to correspond with these properties in that the noun can be in a direct construction
with any cardinal numeral ('two chairs' — only bounded entities can be counted) and
is only used to refer to the whole object (and not to any of its parts).

Table 1. The distribution of three noun categories (whose members are used to refer to a single, concrete object) in a representative sample of the world's languages: SINGULAR OBJECT NOUN, SET NOUN, AND SORT NOUN (Rijkhoff 2004: 30–31).

	SINGULAR OBJECT N^N	SET N^W	SORT N^W
TYPE 3: V-N-A			
Abkhaz	+	+?[−Human/+Animate]	
Alamblak			
Basque		+	
Berbice Dutch		+	
Bukiyip	+		
Burushaski	+	+	
Chukchi		+[NUM=bound, except if case=ABS]	
Dutch	+		
Georgian		+	
Guaraní			
Hittite	+[Masc/Fem. gender]	+[Neuter gender]	
Hungarian		+	
Ika		+	
Kayardild		+	
Ket	+	+	
Hmong Njua		+	[see Rijkhoff 2000]
Nama Hottentot			
Nasioi	+[+Human]	+[−Human]	
Ngalakan		+	
Ngiti	+[+Human]	+[−Human]	
Sumerian		+	
Wambon		+	
TYPE 3/4: V-N(-A)			
Babungo	+		
Bambara			
Gude		+	
Kisi	+		
Oromo		+	
Pipil	+?	+?	
Sarcee	+?[+Human]	+[−Human]	
Tamil	+[+Human]	+[−Human]	
TYPE 4: V-N			
Burmese			+
Galela		+	
Hixkaryana		+	
Koasati			
Korean			+
Krongo	+	+	
Lango		+	
Mandarin Chinese			+
Nivkh		+[NUM=bound]	+
Nung			+
Nunggubuyu	+[+Human]	+	
Tsou		+	
Vietnamese			+
West Greenlandic	+		
Type unknown:			
Nahali		+	

Set nouns, sort nouns and general nouns (all members of flexible or 'underspecified' noun categories) can be said to be semantically vague, with a wide semantic content: values for the features Shape and Homogeneity are either left unspecified (*Homogeneity*) or they are specified in such a way that they do not quite match the properties of a real world concrete object. Furthermore set nouns, sort nouns and general nouns are transnumeral and in the case of general nouns the meaning is so vague it can even be used to refer to, for example, a mass or a collective (see (11) above).

In spite of these semantic differences, singular object nouns, set nouns, sort nouns and general nouns are all equally good exemplars of the lexical category Noun from a cross-linguistic perspective. There are no good reasons to assume that, for example, singular object nouns in English are on the whole better or worse than sort nouns in Thai. The narrow meaning of a singular object noun may seem attractive for those who prefer a close correspondence between the intension and the extension of a noun, but it makes the singular object noun rather inflexible for referential purposes. Conversely, the ambiguous or vague semantics of nouns with a wide meaning (set nouns, sort nouns and general nouns) may be difficult to handle for grammatical theories that do not take linguistic or extra-linguistic contextual factors into account, but it makes them rather versatile ('flexible') for referential purposes (on the role of various kinds of contexts in Functional Discourse Grammar, see Rijkhoff 2008).

4.3 Coercion of flexible lexemes

We have argued in Section 3 (referring to Hengeveld and Rijkhoff 2005) that a vague lexeme like a contentive or a non-verb gets its categorical flavour by placing it in a certain syntactic slot (*syntactic coercion*) or by providing it with a special morphological marker (*morphological coercion*). In the case of nouns that are characterized by vague semantics (set nouns, sort nouns, general nouns), syntactic coercion cannot produce the intended meaning, since (at least in current theories of grammar) the syntactic slot does not distinguish between members of different noun categories.[8] From a morphosyntactic perspective then, the only alternative is coercion by morphological markers, and this is exactly what we find for each of the three noun categories whose members are characterized by vague semantics. Thus we see that singulative or collective markers (dubbed 'nominal aspect markers' in Rijkhoff 2004: Chapter 4) are employed with set nouns, and classifiers (or individualizers) in the case of nouns belonging to the −Shape categories:

(12) <u>Nominal category</u>   <u>Specific sense coerced by</u>
      set noun: singulative or collective marker
      sort noun: sortal classifier[9]
      general noun: general classifier

In Oromo, which has set nouns, the singulative and collective marker specify *what kind of set* is being referred to by the speaker: a singleton set or a collective set (the *kind* distinction is a qualitative rather than a quantitative distinction; Rijkhoff 2004; see Rijkhoff forthcoming for a somewhat different and more recent account in terms of the layered representation of linguistic structures):

> Oromo (Stroomer 1987: 76–77, 84–85)

(13)   Unmarked set: *nama* 'man/men'   vs.   Singleton set: *namica* 'a/the man'

(14)   Unmarked set: *saree* 'dog/dogs'   vs.   Collective set: *sareellee* 'dogs'

In the following examples from Mandarin Chinese, "the classifier *duo* refers to 'bud' of a plant or flower, while *zhu* refers to the plant itself" (Huang and Ahrens 2003: 360). In either case, the classifier coerces the noun to have a certain reading depending on the information 'entailed' in the classifier:

> Mandarin Chinese (Huang and Ahrens 2003: 361 — no tones indicated in original)[10]

(15)   a.   yi   duo   hua   b.   yi   zhu   hua
        one   CL   flower/plant   one   CL   flower/plant
        'one flower'   'one plant'

Thus, classifiers allow the speaker to emphasize different aspects of an entity (Adams 1989: 3). As in the case of contentives and other flexible word classes in Figure 3, certain meaning components of nouns belonging to the flexible category N^W are highlighted whereas others are downplayed. This is illustrated in (16) with the Yucatec examples given in (11).

(16)   *Highlighting/Downplaying* of meaning components of a general noun
        (A B C etc. are the meaning components of Yucatec *háʾas*):

a/one-CLF banana	A	B	C	D	E …	Highlighted properties of *háʾas*
'un-tzʾíit háʾas	+		+		+	A C E … => a banana
'un-wáal háʾas	+	+	+			A B C … => a leaf (of a banana tree)
'un-kúuch háʾas		+		+	+	B D E … => a bunch of bananas
'um-pʾíit háʾas	+		+	+		A C D … => a portion (of a banana mass)

In sum, both in the case of (a) contentives and non-verbs and (b) certain noun categories (set nouns, sort nouns, general nouns) we are dealing with lexemes with a wide semantic content. The actual meaning of contentives and non-verbs in an utterance can be coerced by syntactic or morphological means (Hengeveld et al. 2004), whereas in the case of flexible noun categories a more specific sense is coerced by morphological elements (nominal aspect marker, classifier).

5.   Flexible word classes as true categories

A flexible word class is sometimes regarded as merger of two or more rigid classes (see e.g. Evans and Osada 2005: 366), but this is not correct.[11] A flexible word class like *Contentive* is not some kind of union of the rigid categories Verb, Noun, and Adjective, but a distinct category in itself, just like a rigid word class. Perhaps the main semantic difference between rigid and flexible word classes has to do with the value for certain meaning components. We will see below, for example, that contentives are not characterized by a positive value for the feature Transitive. Thus, the fact that *contentives* can be used in verbal, nominal and adjectival function does not imply that they combine typical properties of verb, nouns and adjectives. Rather, a contentive is neither a verb nor a noun nor an adjective, precisely because it does NOT have certain values for important features (more on this in Section 6 below).

   If a flexible word class it is not the result of some kind of merger of rigid word classes but constitutes a proper lexical category in its own right, then we may expect to find, for instance, that problems regarding category membership of rigid word classes are also relevant for flexible word classes. Such problems are discussed in a recent article by Aarts (2004a), who proposes that varying degrees of category membership and overlaps between categories can be modeled in terms of gradience. He distinguishes between two types of gradience (Aarts 2004a: 1).[12]

> "One type I will call *Subsective Gradience* (SG). It is intra-categorial in nature, and allows for members of a class to display the properties of that class to varying degrees. The other type is called *Intersective Gradience* (IG). This is an inter-categorial phenomenon which is characterized by two form classes 'converging' on each other."

Subsective gradience can be investigated in terms of prototype theory (which says that some members of a category are better examples of the category than others in they resemble the prototype more closely; Rosch 1973), whereas intersective gradience involves cases of categorial overlap, as when a member displays features of more than one category. The difference between subsective and intersective gradience (involving members of the rigid categories Noun and Verb in English) is shown in examples (17) and (18):

> Subsective Gradience: e.g. degree of nouniness (PROTOTYPE THEORY)
> (17) *chair*, *spoon*, …   vs.   *sheep*, *water*, *linguistics*, …
>   more typical   vs.   less typical

In English the prototypical noun is generally assumed to denote a concrete entity (e.g. 'chair') and to inflect for plural number ('chair-s'). In this view, a noun such as 'sheep' is rather atypical (for English, that is) in that it does not inflect for number. In other words, in English 'chair$_N$' is a better examplar of the category Noun than 'sheep$_N$'.

Subsective gradience is also found among members of flexible word classes. Proto-typical properties are known to correlate strongly with certain important aspects of human behaviour and together these correlations are referred to as 'prototype effects' (Cruse 2004: 130–1). One of the prototype effects concerns frequency. What we see in a language such as Samoan is that members of the flexible word class Contentive can be ranked according to frequency of occurrence in certain functions. For example, the more typical exemplars of the category Contentive occur with the same frequency in verbal and nominal function, whereas the less typical exemplars are attested more often in one of these functions (Mosel and Hovdhaugen 1992: 77; see also the quota-tion in note 1).

In the case of intersective gradience we are dealing with vagueness across category boundaries.

> Intersective Gradience: e.g. nominal and verbal properties of gerunds
> (OVERLAP)

(18) a. Brown's PAINTING of his daughter   vs.
          more nominal: 'verbal noun'
      b. Brown's PAINTING his daughter
          more verbal: 'nominal verb'

Thus 'painting' in both (18a) and (18b) has properties of verbs as well as nouns (albeit in different degrees; for details see Aarts 2004a). This kind of gradience, which involves category overlap, is also attested among members of flexible word classes. For example, it has been established for a wide variety of languages that collective markers are a common source for plural markers. This diachronic development from collective to plural marker has been observed in, for example, the Kartvelian, the Mesoamerican and the Semitic languages (Rijkhoff 2004: 117). Since collective markers appear with set nouns and plural marking is a property of singular object nouns, we may hypoth-esize that, along with the change from collective to plural marker, set nouns have be-come singular object nouns in the languages in question. Since such a change does not happen overnight, we may assume that for some time the old category of set nouns must have overlapped partially with the new, emerging category of singular object nouns in the languages where this change has occurred.

6.   Flexible and rigid nouns and grammar

The distinction between flexible and rigid nouns is important for various reasons. First of all, it improves our general understanding of PoS systems in that it adds a new dimension to current classifications of lexical word categories (Figure 3). Secondly, the features Shape and Homogeneity, which have been used to distinguish between the various noun categories (Figure 4), seem to correlate with important grammatical

phenomena such as number agreement, more specifically with so-called number discord. Consider the following examples:

Georgian (Harris 1981: 22)
(19) sami ḳnuṭi goravs
      three kitten roll:3sg
      'Three kittens are rolling.'

Lango (Noonan 1992: 168)
(20) gúlú àdêk ɔ̀tɔ̀ɔ̀
      pot   three 3sg:die:perf
      'Three pots broke.'

Oromo (Stroomer 1987: 107)
(21) gaala lamaani sookoo ḍak'-e
      camel two   market go-3sg.m.past
      'Two camels went to the market.'

In each example we see that a plural subject triggers singular agreement with the person/number marker on the verb. In my sample (Figure 2) this phenomenon is only attested in languages with set nouns, a flexible noun category whose members denote spatially bounded properties but are neutral with respect to the feature Homogeneity (Rijkhoff 2004: 105). Apparently, agreement in these languages is with the set, which — irrespective of the number of individuals it contains — is always a singular entity (Rijkhoff 1993).[13] In other words, it seems we can only make sense of so-called number discord (but notice that the term appears to be a misnomer) if we take into account the fact that different languages use different noun categories (some of them flexible) to refer to concrete objects in the physical world.

The features that we used to specify the difference between rigid and flexible noun categories also play a role in a recent attempt to account for the PoS hierarchy (Rijkhoff 2003).

Parts of Speech hierarchy (Hengeveld 1992)
(22) Verb > Noun > Adjective > (manner) Adverb

This hierarchy says that, if a languages employs members of a certain rigid category 'down' the hierarchy (e.g. Adjective), it also employ members of the rigid categories 'up' the hierarchy (Verb, Noun). One question that could be asked regarding this hierarchy is: what are the necessary conditions to allow for appearance of the next rigid word class in the hierarchy? For example, some languages with the category Noun also have the category Adjective (languages of Type 3 in Figure 3), but other languages do not (languages of Type 4 in Figure 3). It appears to be the case that adjectives are only attested in languages with +Shape noun categories (Singular Object Noun, Set Noun; see *Table* 1 above).[14] This is also seen in Figure 5, which shows that there are no lan-

✓ *If a language has adjectives, it has +Shape nouns*

WHEN CAN A LANGUAGE HAVE ADJECTIVES?	LANGUAGES WITH A MAJOR, DISTINCT CLASS OF ADJECTIVES (TYPE 3 IN FIGURE 3)	LANGUAGES WITHOUT A MAJOR, DISTINCT CLASS OF ADJECTIVES
LANGUAGES WITH +SHAPE NOUNS	Type 3. Abkhaz, Alamblak, Basque, Berbice Dutch Creole, Bukiyip (=Mountain Arapesh), Burushaski, Dutch, Georgian, Guaraní, †Hittite, **Hmong Njua (see Rijkhoff 2000)**, Hungarian, Ika, Kayardild, Ket, Nama Hottentot, Nasioi, Ngalakan, Ngiti, Oromo, †Sumerian, Wambon	Type 3/4: Babungo, Bambara, Chukchi, Gude, Kisi, Oromo, Pipil, Sarcee, Tamil Type 4: Galela, Hixkaryana, Koasati, Krongo, Lango, Nunggubuyu, Tsou, West Greenlandic Type 4/5. Cayuga
LANGUAGES WITH −SHAPE NOUNS	**(no languages)**	Type 4. Burmese, Korean, Mandarin Chinese, Nivkh, Nung, Vietnamese
?	†Etruscan, †Meroitic, Nahali	
FLEXIBLE		*Samoan, †Hurrian, Imbabura Quechua, Turkish*

Figure 5. Languages with and without a major, distinct class of adjectives.

guages with adjectives that employ −Shape nouns to refer to a single, concrete object (Rijkhoff 2000). Apparently having +Shape nouns is a necessary (though not a sufficient) condition for a language to have a distinct class of adjectives. In other words, if a language has adjectives, then it uses +Shape nouns (i.e. individual object nouns or set nouns) to refer to a single concrete object (Figure 6). From this we may conclude that nouns in a language must be defined 'narrowly' with respect to the feature Shape before it can have a distinct class of adjectives.

Further research suggests that there are also certain conditions that must be met before a language can have the categories Verb or Noun in its PoS system (Rijkhoff 2003). Figure 7 shows that languages with the category Contentive (Type 1 in Figure 3) are the only languages without transitive lexemes in the basic lexicon.[15]

Transitive lexemes designate a dynamic relationship between two obligatory participants: an agent/subject and a patient/object. Whereas all languages with a major,

Flexible	1	Contentive		
	2	Verb		Non-verb
Rigid	3	**Verb**	**Noun [+Shape]**	**Adjective**
	4	Verb		Noun [±Shape]
	5	Verb		

Figure 6. Modified classification of PoS systems: a language only has the category Adjective, if it uses members of a +Shape noun category to refer to a single, concrete object.

✓ If a language has nouns, it has transitive lexeme (i.e. verbs) — but not the other way around.

WHEN CAN A LANGUAGE HAVE NOUNS?	LANGUAGES WITH A MAJOR, DISTINCT CLASS OF NOUNS (TYPES 3, 3/4, 4 IN FIGURE 3)	LANGUAGES WITHOUT A MAJOR, DISTINCT CLASS OF NOUNS
LANGUAGES WITH TRANSITIVE LEXEMES	Type 3: Abkhaz, Alamblak, Basque, Berbice Dutch Creole, Bukiyip (=Mountain Arapesh), Burushaski, Dutch, Georgian, Guaraní, †Hittite, Hmong Njua, Hungarian, Ika, Kayardild, Ket, Nama Hottentot, Nasioi, Ngalakan, Ngiti, †Sumerian, Wambon Type 3/4: Babungo, Bambara, Chukchi, Gude, Kisi, Oromo, Pipil, Sarcee, Tamil Type 4: Burmese, Galela, Hixkaryana, Koasati, Korean, Krongo, Lango, Mandarin Chinese, Nivkh, Nung, Nunggubuyu, Tsou, Vietnamese, West Greenlandic	Type 2 [non-verbs]: †Hurrian, Imbabura Quechua, Turkish Type 4/5[minor N]: Cayuga
LANGUAGES WITHOUT TRANSITIVE LEXEMES	(no languages)	Type 1: Samoan (*outside this sample e.g. Halkomelem, Squamish, and other Salish lgs.*)
?	†Etruscan, †Meroitic, Nahali	

Figure 7. Languages with and without a major, distinct class of nouns.

distinct class of verbs have a set of basic transitive lexemes in the lexicon, such lexemes are absent in Samoan (Rijkhoff 2003):

> "With the exception of a very small class of locative verbs […], Samoan verbs do not require more than one argument, i.e. S or O. If we define obligatory transitive verbs as bivalent verbs which express transitive actions and which require two arguments referring to the agent and the patient, then Samoan does not have obligatory transitive verbs." (Mosel 1991: 188)

> "If we compare Samoan verbs with transitive and intransitive verbs in other languages where these two categories are distinguished in terms of the number of obligatory arguments, then there are no cardinal transitive verbs in Samoan, i.e. bivalent verbs expressing transitive actions. Except for a very small class […], all Samoan verbs (including ergative verbs) maximally require one argument, namely S or O, both of which are expressed by absolutive noun phrases in basic verbal clauses." (Mosel & Hovdhaugen 1992: 724)

✓ *If a language has verbs, it has transitive lexemes (and vice versa).*

WHEN CAN A LANGUAGE HAVE VEBRS?	LANGUAGES WITH VERBS (TYPES 2, 3, 3/4, 4, 4/5)	LANGUAGES WITHOUT A MAJOR, DISTINCT CLASS OF VERBS (TYPE 1)
LANGUAGES WITH TRANSITIVE LEXEMES	Type 2: †Hurrian, Imbabura Quechua, Turkish Type 3: Abkhaz, Alamblak, Basque, Berbice Dutch Creole, Bukiyip (=Mountain Arapesh), Burushaski, Dutch, Georgian, Guaraní, †Hittite, Hmong Njua, Hungarian, Ika, Kayardild, Ket, Nama Hottentot, Nasioi, Ngalakan, Ngiti, Oromo, †Sumerian, Wambon Type 3/4: Babungo, Bambara, Chukchi, Gude, Kisi, Oromo, Pipil, Sarcee, Tamil Type 4: Burmese, Galela, Hixkaryana, Koasati, Korean, Krongo, Lango, Mandarin Chinese, Nivkh, Nung, Nunggubuyu, Tsou, Vietnamese, West Greenlandic Type 4/5: Cayuga	(no languages)
LANGUAGES WITHOUT TRANSITIVE LEXEMES	(no languages)	Type 1 [Contentive]: Samoan (*outside this sample e.g. Halkomelem, Squamish, and other Salish languages*)
?	†Etruscan, †Meroitic, Nahali	

Figure 8. Languages with and without a major, distinct class of verbs (see note 15).

Having a set of transitive lexemes is only a necessary condition for a language to have a category Noun as there are also languages with transitive lexemes but without nouns (e.g. languages of Type 2 in Figure 3), but having transitive lexemes is both a necessary and a sufficient condition for a language to have the category Verb (Figure 8).

These data from a representative sample of the world's languages indicate that the number of distinct word classes in a language (and the order in which they appear) is determined by the value for certain lexical features (TRANSITIVE, SHAPE). Only when the relevant feature has a positive value (+TRANSITIVE, +SHAPE) does the PoS system allow for the occurrence of the next word class in the hierarchy.

(23)   Features: **necessary** conditions in the PoS hierarchy (Rijkhoff 2003):

VERB   →   NOUN   →   ADJECTIVE   →   MANNER ADVERBS
[+TRANSITIVE]   [+SHAPE]   [+GRADABLE?]

Recall that the presence of a set of transitive items in the basic lexicon is a necessary and sufficient condition for a language to have a distinct class of verbs, but only a necessary condition for the occurrence of a distinct class of nouns in the PoS of a language. The adjectival feature Gradable has a question mark, because it still has to be established whether it is a positive value for this particular feature that makes it possible for manner adverbs to occur in the PoS system of a language.

At a more abstract level these data seem to suggest that a language can only have distinct classes of verbs, nouns, and adjectives if the basic meaning of lexical items somehow encodes the prototypical properties of temporal and spatial entities (events and things). The prototypical event is an activity that involves an agent and a patient (+Transitive); the prototypical thing is a concrete object (+Shape). Thus, a language can only have major, distinct classes of verbs, nouns and adjectives if the lexicon contains (a) items that designate a dynamic relationship between an agent and a patient, and (b) items that designate a property that is specified as having a boundary in the spatial dimension.

7.   Conclusion

In this article I have tried to show that, across languages, the rigid category Noun itself includes rigid and flexible noun types. It appears that members of four noun categories are used by speakers of different languages when they talk about a single, concrete object in the external world: Singular Object Noun, Set Noun, Sort Noun and General Noun. Singular object nouns (+Shape, −Homogeneity) can be called rigid: they are characterized by a narrow semantic content in that the values for the features Shape and Homogeneity match properties of the kind of entity they denote (a single, concrete object). As in the case of the flexible parts of speech Contentive and Non-verb, members of the flexible noun categories Set Noun, Sort Noun and General Noun have vague semantics (i.e. they have a wide semantic content). Flexible nouns have a neutral value for the feature Homogeneity and/or a negative value for the lexical feature Shape, which does not correspond with the fact that a concrete object in the external word is typically a non-homogeneous entity with a definite spatial contour. Individualizing strategies are required to make the property denoting members of the −Shape noun categories Sort Noun and General noun countable (Section 4).

I then argued that a flexible word category is not a merger of some rigid word classes, as has sometimes been suggested. A flexible word class is a true category, whose members can be ranked in terms of category membership, just like the members of a rigid word class. Finally I showed how the distinction between flexible and rigid noun

categories provides new insights into PoS systems of languages spoken across the globe and how these systems interact with other parts of the grammar. Last but not least, I have claimed that the features that are relevant for the distinction between flexible and rigid categories (Transitive, Shape) also play a central role in an account of the parts of speech hierarchy.

Notes

1. The text continues as follows (Mosel and Hovdhaugen 1992: 77): "Not all roots occur with the same frequency as verbs and nouns. Some roots predominantly function as verbs, whereas others are more likely to be found in the function of nouns. Until now we have not, for instance, found *alu* "go" in a nominal function [but, as was noted in Hengeveld and Rijkhoff (2005: 412), the authors actually provided an example of *alu* in nominal function themselves some pages earlier — JR] or *mea* "thing" in a verbal function […]. But we hesitate to say that *alu* is inherently a verb and mea inherently a noun for two reasons. Firstly, we cannot find any functional explanation why *alu* should not be used as a noun and mea as a verb, whereas, for instance, *gaoi* "thief, to steal" and *tagata* "person, to be a person" are bifunctional. And, secondly, previous experience taught us to be careful with classifications. The more texts we analyzed, and included in our corpus, the more items were unexpectedly found in nominal or verbal function."

2. When a cardinal numeral modifies a set noun, it specifies the size of the set (i.e. the number of individuals contained in the set) rather than the number of sets (Rijkhoff 2004: 147).

3. Sort nouns and general nouns are both vague in the sense that their meaning definition does not seem to contain any information about the spatial contour of the entity. This does not necessarily mean that the speakers of e.g. Thai or Yucatec do not *know* that a table in the physical world is a discrete object, but rather that this particular piece of knowledge is simply not part of the meaning of the noun (instead it would be part of e.g. the 'encyclopedic knowledge' about the referent of the NP; for a similar point see e.g. Unterbeck 1993). However, the idea that the employment of members of different nominal categories (*Seinsarten*) for the same thing may reflect "differences in the ontological beliefs the speakers of these languages hold about the referents of nouns" is discussed in e.g. Foley (1997: 231). On the subject of *ontological relativity* see also Quine (1960, 1969), Lucy (1992), Imai and Gentner (1993).

4. Since this is not an ontological but a linguistic classification, there is in principle no direct relationship between noun category (*Seinsart*) and real-world entity (*Sein*-correlate). Hence the same object can be referred to by using nouns from four different categories.

5. I am not aware of other classifications of nouns based on facts from a wide range of typologically different languages. Consequently this classification differs in some important respects from other, non-typological discussions of noun categories (for example, it contains some novel categories: set nouns, general nouns). Furthermore this classification does not assign any special status to plural nouns (cf. Jackendoff 1991), simply because plural nouns do not represent a different noun category (categorization is about types not tokens). Notice, incidentally, that boundaries between noun categories are not always clear cut. For example, simple tools can be more or less homogeneous objects (if you break a stick you have two sticks), i.e. certain concrete objects

are +Shape and +Homogenous but they are not collectives. Categorization is also a matter of (inadequate) perception (Russell 1923; cf. Aarts 2004a: 27; Langacker 2006: 116). Thus 'hair' is a mass noun in 'much hair' but an individual object noun in 'many hairs'. Finally, category membership may not always be clear due to the gradual nature of language change (on the change from one noun category to another, see Rijkhoff 2004: 117).

6. It is not clear why the feature Shape is relevant for all *Seinsarten*, whereas the feature Homogeneity helps to define only four nominal subcategories. Yet this classification confirms Friedrich's observation that "the category of shape appears to be a typological universal in grammar [..], and of not inconsiderable significance for a theory of semantics in grammar" (Friedrich 1970: 380; see also Wierzbicka 2006). The fact that in this classification the feature Shape is more important than the feature Homogeneity (Figure 4) may also have to do with the idea that spatial orientation is primary in human cognition and links up nicely with observations in other grammatical domains ('Localism').

7. Speakers of a language may use more than one of these noun types to refer to concrete objects. In Ngiti, for example, singular object nouns are used for human entities and set nouns for nonhuman entities (Kutsch Lojenga 1994: 340, 355).

8. But one can easily imagine a theory employing NP structures with syntactic slots that can only be filled by members of certain noun categories (Singular Object Noun, Mass noun, etc.).

9. For a recent treatment of classifiers in Mandarin Chinese see Huang and Ahrens (2003: 360), who argue that "contextual information will 'coerce' a sense by eliminating other possible interpretations in a richly encoded but under-specified lexicon". Interestingly, they add that semantic coercion can be predicted through a well encoded qualia structure as proposed in Pustejovsky (1995).

10. The situation may be more complex than what is shown here. At least some Mandarin nouns seem to be able to combine with the three different types of classifiers distinguished by Huang and Ahrens (2003), each type coercing another sense of the same noun. INDIVIDUAL CLASSIFIERS (CLF_IND) are attested most often and are used to refer to a concrete object; KIND CLASSIFIERS (CLF_KD) are used to refer to a particular kind of entity, and EVENT CLASSIFIERS (CLF_EV) coerce an event reading (Huang and Ahrens 2003: 371)

Individual	Kind	Event
a. yi jia feiji	b. yi zhong feiji	c. yi ban feiji
one CLF_IND airplane	one CLF_KD airplane	one CLF_EV airplane
'one airplane'	'one kind of airplane'	'one scheduled flight'

These data suggest that at least some Mandarin nouns are perhaps even more vague and consequently even more versatile or flexible than previously assumed (Huang and Ahrens 2003: 355).

11. Allowing for vagueness is not an attractive option for some linguists, perhaps because of the strong human propensity to see the world as being structured in terms of discrete entities (cf. Aarts 2004a: 4; see also Aarts 2004b, Aarts 2006).

12. Ultimately Aarts (2004a) defends a position that allows for gradience, while keeping sharp boundaries between categories. See also Sorace and Keller (2005) on gradience in grammar.

13. Note that in the case of collectives one can also have agreement with the collective (singular) or with the individuals in the collective (plural); see e.g. Bock et al. 2006.

14. Notice that this does not mean that adjectives cannot occur in a classifier language, which typically but not necessarily has Sort Nouns or General Nouns (both of which have a negative value for the feature Shape). In many languages numeral classifiers have developed into markers of other grammatical categories such as definiteness, specificity or topicality (Rijkhoff 2004: 51). In such cases the erstwhile classifiers no longer serve as 'individualizers' in the sense of Lyons (see Section 4.1). For example, nouns in Hmong (in bold print in *Figure* 5) have all the properties of Set Nouns (+Shape) and the element that used to be a classifier now serves as a collective aspect marker (Rijkhoff 2000). Another reason why adjectives and classifiers are not mutually exclusive is that some languages use Singular Object Nouns or Set Nouns (both +Shape) as well as Sort Nouns or General Nouns (both −Shape) to refer to concrete objects. This seems to be the case in Yucatec, which has a group of General Nouns (Section 4.1) but also a group of animate nouns that are probably best categorized as Set Nouns: they may occur with an optional so-called "plural marker" which is absent when the NP contains a numeral (Lucy 1992: 46–49).

15. I use the neutral label 'transitive lexeme' to avoid being forced to discuss lexemes in terms of the traditional word classes (Verb, Noun, etc.), but since having a set of transitive lexemes is both a necessary AND a sufficient condition for a language to have a class of verbs, 'transitive lexeme' is of course synonymous with 'verb' (Figure 8).

References

Aarts, Bas. 2004a. Modelling linguistic gradience. *Studies in Language* 28(1): 1–49.

Aarts, Bas. 2004b. Conceptions of gradience in the history of linguistics. *Language Sciences* 26(4): 343–389.

Aarts, Bas. 2006. Conceptions of categorization in the history of linguistics. *Language Sciences* 28(4): 361–385.

Aarts, Bas; David Denison; Evelien Keizer & Gergana Popova (eds.). 2004. *Fuzzy grammar: A reader*. Oxford: Oxford University Press.

Adams, Karen Lee. 1989. *Systems of numeral classification in the Mon-Khmer, Nicobarese, and Aslian subfamilies of Austroasiatic*. Canberra: Australian National University (Pacific Linguistics B-101).

Andrzejewski, Bogumił W. 1960. The categories of number in noun forms in the Borana dialect of Galla. *Africa* 30: 62–75.

Bisang, Walter. 1996. Areal typology and grammaticalization: processes of grammaticalization based on nouns and verbs in East and mainland South East Asian languages. *Studies in Language* 20(3): 519–597.

Bock, Kathryn; S. Butterfield; Anne Cutler; J. Cooper Cutting; Kathleen M. Eberhard & Karen R. Humphreys. 2006. Number agreement in British and American English: disagreeing to agree collectively. *Language* 82: 64–113.

Bohnemeyer, Jürgen. 2002. *The grammar of time reference in Yucatec Maya*. Munich: LINCOM Europa [LINCOM Studies in Native American Linguistics 44].

Broschart, Jürgen & Carmen Dawuda. 2000. Beyond nouns and verbs: typological studies in lexical categorisation. Working Paper no. 113 of *Arbeiten des Sonderforschungsbereichs 282*

(Düsseldorf — Wuppertal — Köln), Institut für Sprache und Information, Heinrich-Heine-Universität, Düsseldorf.

Cheng, Lisa Lai-Shen & Rint Sybesma. 1999. Bare and not-so-bare nouns and the structure of NP. *Linguistic Inquiry* 30: 509–542.

Czaykowska-Higgins, Ewa & M. Dale Kinkade. 1998. Salish languages and linguistics. *Salish Languages and Linguistics: theoretical and descriptive perspectives*, Ewa Czaykowska-Higgins & M. Dale Kinkade (eds.), 1–68. Berlin: Mouton de Gruyter (Trends in Linguistics: Studies and Monographs 107).

Cruse, D. Alan. 1986. *Lexical semantics*. Cambridge: Cambridge University Press.

Cruse, D. Alan. 2004. *Meaning in language: an introduction to semantics and pragmatics* (2nd edition). Oxford: Oxford University Press.

Evans, Nicholas & Toshiki Osada. 2005. Mundari: the myth of a language without word classes. *Linguistic Typology* 9(3): 351–390.

Foley, William A. 1997. *Anthropological linguistics: an introduction*. Oxford: Blackwell.

Friedrich, Paul. 1970. Shape in grammar. *Language* 46(2): 379–407 [Reprinted in Paul Friedrich, 1979, *Language, context, and the imagination*, 340–359. Stanford, CA: Stanford University Press].

Harris, Alice Carmichael. 1981. *Georgian syntax: a study in relational grammar*. Cambridge: Cambridge University Press (Cambridge Studies in Linguistics 33).

Hengeveld, Kees. 1992. *Non-verbal Predication: theory, typology, diachrony*. Berlin: Mouton de Gruyter.

Hengeveld, Kees; Jan Rijkhoff & Anna Siewierska. 2004. Parts-of-speech systems and word order. *Journal of Linguistics* 40(3): 527–570.

Hengeveld, Kees & Jan Rijkhoff. 2005. Mundari as a flexible language. *Linguistic Typology* 9(3): 406–431.

Huang, Chu-Ren & Kathleen Ahrens. 2003. Individuals, kinds and events: classifier coercion of nouns. *Language Sciences* 25: 353–373.

Hundius, Harald & Ulrike Kölver. 1983. Syntax and semantics of numeral classifiers in Thai. *Studies in Language* 7(2): 164–214.

Ikegami, Y. 1985. 'Activity'-'Accomplishment'-'Achievement': a language that can't say 'I burned it but it didn't burn' and one that can. *Linguistics and philosophy: essays in honor of Rulon S. Wells*, A. Makkai & A. K. Melby (eds.), 265–304. Amsterdam: Benjamins.

Imai, M. & D. Gentner. 1993. Linguistic relativity vs. universal ontology: cross-linguistic studies of the object/substance distinction. *What we think, what we mean and how we say it: papers from the parasession on the correspondence of conceptual, semantic and grammatical representations*, K. Beals; G. Cooke; D. Kathman; S. Kita; K. McCullough & D. Teston (eds.), 171–186. Chicago: Chicago Linguistic Society.

Jackendoff, Ray. 1991. Parts and boundaries. *Cognition* 41: 9–45.

Kuhn, Wilfried. 1982. Kollektiva und die Technik KOLLEKTION am Beispiel des Deutschen. *Apprehension. Das sprachliche Erfassen von Gegenständen. Teil I: Bereich und Ordnung der Phänomene*. Hansjakob Seiler & Christian Lehmann (eds.), 84–97. Tübingen: Gunter Narr (LUS, 1/I).

Langacker, Ronald W. 2006. On the continuous debate about discreteness. *Cognitive Linguistics* 17(1): 107–151.

Lucy, John A. 1992. *Grammatical categories and cognition: a case study of the linguistic relativity hypothesis*. Cambridge: Cambridge University Press [Studies in the Social and Cultural Foundations of Language 13].

Lyons, John. 1977. *Semantics* (2 volumes). Cambridge: Cambridge University Press.

Mithun Williams, Marianne. 1976. *A grammar of Tuscarora*. New York: Garland.

Mosel, Ulrike. 1991. Transitivity and reflexivity in Samoan. *Australian Journal of Linguistics* 11: 175–194.

Mosel, Ulrike and & Hovdhaugen. 1992. *Samoan reference grammar*. Oslo: Universitetsforlaget AS.

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

Pustejovsky, James. 1995. *The generative lexicon*. Cambridge, MA: MIT Press.

Quine, Willard Van Orman. 1960. *Word and object*. Cambridge, MA: MIT Press.

Quine, Willard Van Orman. 1969. *Ontological relativity and other essays*. New York: Columbia University Press.

Rijkhoff, Jan. 1993. 'Number' disagreement. *Proceedings of the XVth International Congress of Linguists*, André Crochetière; Jean-Claude Boulanger & Conrad Ouellon (eds.), 274–276. Sainte-Foy, Québec: Presses de l'Université Laval.

Rijkhoff, Jan. 2000. When can a language have adjectives? An implicational universal. *Approaches to the typology of word classes,* Petra M. Vogel & Bernard Comrie (eds.), 217–257. Berlin/New York: Mouton de Gruyter.

Rijkhoff, Jan. 2003. When can a language have nouns and verbs? *Acta Linguistica Hafniensia* 35: 7–38.

Rijkhoff, Jan. 2004. *The noun phrase*. Oxford: Oxford University Press [expanded paperback edition of 2002 hardback publication].

Rijkhoff, Jan. 2008. Layers, levels and contexts in Functional Discourse Grammar. *The noun phrase in Functional Discourse Grammar*, Daniel García Velasco & Jan Rijkhoff (eds.), 63–115. Berlin/New York: Mouton de Gruyter.

Rosch, Eleanor. 1973. Natural categories. *Cognitive Psychology* 4: 328–350.

Russell, Bertrand. 1923/1996. Vagueness. *Australian Journal of Philosophy and Psychology* 1, 84–92 [Reprinted in Rosanna Keefe & Peter Smith (eds.), 1996, *Vagueness: A reader*, 61–68. Cambridge, MA: MIT Press; also reprinted in Aarts *et al.* 2004.]

Sasse, Hans-Jürgen. 1993. Syntactic categories and subcategories. *Syntax: an international handbook of contemporary research* (2 vols.), Joachim Jacobs; Arnim vom Stechow; Wolfgang Sternefeld & Theo Vennemann (eds.), 646–686. Berlin: Walter de Gruyter.

Schachter, Paul. 1985. Parts-of-speech systems. *Language typology and syntactic description. Volume I: Clause structure*, Timothy Shopen (ed.), 3–61. Cambridge: Cambridge University Press.

Smit, Niels. 2001. *De rol van derivatie bij lexicale specialisatie*. MA thesis, Department of Linguistics, University of Amsterdam.

Song, Jae Jung. 2001. *Linguistic typology: morphology and syntax*. Harlow: Pearson [Longmans Linguistics Library].

Sorace, Antonella & Frank Keller. 2005. Gradience in linguistic data. *Lingua* 115: 1497–1524.

Stroomer, Harry. 1987. *A comparative study of three southern Oromo dialects in Kenya: phonology, morphology and vocabulary*. Hamburg: Buske [Cushitic Language Studies 6].

Unterbeck, Barbara. 1993. *Kollektion, Numeralklassifikation und Transnumerus*. Frankfurt am Main: Peter Lang.

van Baarda, M.J. 1908. *Leiddraad bij het bestuderen van 't Galela'sch dialekt, op het eiland Halmaheira* [Manual for the study of the Galela dialect, on the island of Halmahera]. The Hague: Nijhoff.

Wierzbicka, Anna. 2006. Shape in grammar revisited. *Studies in Language* 30(1): 115–177.

Wilkins, David P. 2000. Ants, ancestors and medicine: a semantic and pragmatic account of classifier constructions in Arrernte (Central Australia). *Systems of nominal classification*, Gunter Senft (ed.), 147–216. Cambridge: Cambridge University Press [Language, Culture and Cognition 4].

Author's address

Jan Rijkhoff
Institut for Lingvistik, Aarhus Universitet
Nordre Ringgade 1, Building 1410, Office 338
8000 Aarhus C
Denmark

linjr@hum.au.dk