

John Benjamins Publishing Company



This is a contribution from *Functions of Language 19:1*
© 2012. John Benjamins Publishing Company

This electronic file may not be altered in any way.

The author(s) of this article is/are permitted to use this PDF file to generate printed copies to be used by way of offprints, for their personal use only.

Permission is granted by the publishers to post this file on a closed server which is accessible to members (students and staff) only of the author's/s' institute, it is not permitted to post this PDF on the open internet.

For any other use of this material prior written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

Please contact rights@benjamins.nl or consult our website: www.benjamins.com

Tables of Contents, abstracts and guidelines are available at www.benjamins.com

Semantic categories in the indigenous languages of Brazil*

Kees Hengeveld, Maria Luiza Braga*, Elisiene de Melo Barbosa*, Jaqueline Silveira Coriolano*, Juliana Jezuiño da Costa*, Mariana de Souza Martins*, Diego Leite de Oliveira*, Vinicius Maciel de Oliveira*, Luana Gomes Pereira*, Liliane Santana**, Cassiano Luiz do Carmo Santos*, Viviane dos Ramos Soares*

University of Amsterdam / Universidade Federal do Rio de Janeiro* /
Universidade Estadual Paulista**

This article investigates which semantic categories, as defined in Functional Discourse Grammar, formally manifest themselves in a sample of native languages of Brazil, and the extent to which the distribution of these manifestations across categories can be described systematically in terms of implicational hierarchies. The areas subjected to investigation are basic interrogative words, basic demonstrative words, and nominalization strategies.

1. Introduction

The aim of this study is to uncover the extent to which a range of semantic categories is relevant to the grammar of indigenous languages of Brazil. Semantic categories are defined in terms of the framework of Functional Discourse Grammar (FDG) and will be introduced in Section 2. The sample used for this study is introduced in Section 3. After that, we will address our general research question, by investigating three grammatical domains: (i) basic question words in Section 4, (ii) basic demonstrative words in Section 5; and (iii) nominalizations in Section 6. In each case, we will capture the results of our study in terms of an implicational hierarchy. In the concluding Section 7 we compare the results of the three partial studies in the light of our general research question.

2. Semantic categorization in FDG

Functional Discourse Grammar (FDG, Hengeveld and Mackenzie 2008) is a typologically-oriented approach to language structure, and distinguishes an Interpersonal, a Representational, a Morphosyntactic, and a Phonological Level of linguistic organization. Semantics is captured in FDG at the Representational Level. In building up the Representational Level, use is made of representational frames, which have a layered, hierarchical structure, and are constructed in a stepwise manner, starting with the hierarchically highest layer and ending with the lowest ones (e.g. Hengeveld & Smit 2009). Every layer has a variable the identity of which corresponds with the semantic category designated by that layer. A preliminary inventory of these, taken from Hengeveld & Mackenzie (2008), is given in (1):

- (1) p *Propositional Content*: a mental construct, locatable in neither space nor time
- ep *Episode*: a thematically coherent combination of States-of-Affairs that is characterized by unity of Time, Location, and Individuals
- e *State-of-Affairs*: an event or state, locatable in time
- f *Property*: a quality or a relation, without independent existence, but applicable to other semantic categories
- x *Individual*: an animate or inanimate object, locatable in space
- l *Location*: a place where an Episode or a State-of-Affairs takes place or where an Individual is located
- t *Time*: a period during which an Episode or a State-of-Affairs takes place
- m *Manner*: a way in which a State-of-Affairs is carried out
- q *Quantity*: an amount or number of Individuals or States-of-Affairs
- r *Reason*: a cause or reason for a State-of-Affairs to occur

Table 1, adapted from Hengeveld & Mackenzie (2008:136), illustrates these categories using representative lexical items from English.

Though we may find lexical representatives of this type in all languages, not all semantic categories can be claimed to be relevant to the grammar of all languages. Only when some aspect of the grammar of a certain language is sensitive to semantic categorization are semantic categories assumed to be relevant to the analysis of that language. While the first five categories in the list above (p, ep, e, f, and x) are probably universally relevant, the latter five may manifest themselves to varying degrees in the languages of the world. This means that it is a matter of empirical research to establish which semantic categories are relevant for the grammar of which languages, on the basis of which one may then try to come to higher-level conclusions concerning the typological distribution of sensitivity to semantic categorization across languages.

Table 1. Semantic categories illustrated by means of English lexemes

Description	Variable	Example
Propositional Content	p	<i>idea</i>
Episode	ep	<i>incident</i>
State-of-affairs	e	<i>meeting</i>
Property	f	<i>colour</i>
Individual	x	<i>chair</i>
Location	l	<i>top</i>
Time	t	<i>week</i>
Manner	m	<i>way</i>
Quantity	q	<i>litre</i>
Reason	r	<i>reason</i>

A straightforward example of the sensitivity of grammars to the semantic categories listed in (1) concerns the extent to which languages distinguish these in processes of nominalization (see Hengeveld & Mackenzie 2008: 131f). Consider some nominalization procedures that English has available for part of the semantic categories listed in (1). These are:

- (2) p: $-\emptyset$: *hope- \emptyset , belief- \emptyset*
 ep: $-\emptyset$: *end- \emptyset , start- \emptyset*
 e: *-tion* *explora-tion, deple-tion*
-age *break-age, cover-age*
 f: *-ness* *mean-ness, kind-ness*
-ity *elastic-ity, rapid-ity*
 x: *-er*: *writ-er, sing-er*
-ant: *inhabit-ant, contest-ant*
 l: *-ery*: *bak-ery, brew-ery*

English has nominalization strategies for propositional contents (p), episodes (ep), states-of-affairs (e), properties (f), individuals (x), and locations (l), but not for times (t), manners (m), quantities (q), and reasons (r). There are, however, other languages that do have nominalization strategies for these latter categories. Supyire (Carlson 1994: 113, 114, 548) for instance, has distinct time, manner, and reason nominalizations, as illustrated in (3)–(5) respectively:

- Supyire
 (3) *Utèè-kwuu-ní nyε à mɔ mé.*
 HISTEMPNR-die-DEF.CL NEG PRF be.long.time NEG
 ‘The time of his death was not long ago.’ (Carlson 1994: 113)

- (4) *Pyii-bíí sàhà nyε na byíí pi tanjáà*
 children-DEF not.yet be PROG raise.IMPf their yesterday
byí-ηká-ni na mé.
 raise-MANN.NR-DEF.CL ON NEG
 ‘Children are no longer raised the way they were raised in the past.’
 (Carlson 1994: 114)
- (5) *Sànyi kà-wyiini li nyε*
 death.announcement.DEF REASNR-announce.DEF(CL.SG) it(CL.SG) COP
pùcèribílá à ñdiré ye?
 female.clan.member.DEF to which.EMPH(CL.SG) Q
 ‘What is the reason for announcing the death to the female clan members?’
 (Carlson 1994: 548)

No examples of quantity nominalizations are known to us.

This paper studies the crosslinguistic availability of distinct grammatical means to express semantic categories with respect to three different phenomena that may show sensitivity to semantic categorization: basic interrogative words, basic demonstrative words, and nominalizations. These topics have each been addressed separately in the FDG literature. Hengeveld & Mackenzie (2008), touched upon above, discuss nominalizations in some detail, Mackenzie (2009) studies the distribution of basic question words in a world-wide sample of 50 languages, and George (2010) studies the distribution of basic demonstrative words in that same sample. We will come back to these studies in the relevant sections. Here it may suffice to say that the current study differs from the ones just mentioned in that it tries to generalize across the three phenomena, and in that it studies languages from one specific area, Brazil. The reason for restricting the study to a single area is that the earlier studies arrive at statistically relevant but not at absolute generalizations as regards the ranking of semantic categories, which may reflect areal differences. By restricting ourselves to a single area that is rich in terms of the different genealogies of the languages involved, we hope to be able to control for this.

Like Mackenzie (2009) we will limit ourselves to the categories Individual (x), Location (l), Time (t), Manner (m), Quantity (q), and Reason (r). The reason for this is that the data available from descriptive grammars are generally insufficient to come to conclusions about the differentiation between Propositional Contents (p), Episodes (ep), States-of-Affairs (e), and Properties (f). The demonstrative *that* in English, for instance, may be used to refer to all of these as well as to inanimate Individuals (x), but this is a topic that is to our knowledge not explicitly addressed in any standard grammatical description of English. This is the more true for languages with a much more limited tradition of grammatical description.

Table 2. Semantic categorization in English

		basic interrogative words	basic demonstrative words	nominalizations
x	Individual	who, what	this, that	brew-er
l	Location	where	there, here	brew-ery
t	Time	when	then, now	–
m	Manner	how	thus	–
r	Reason	why	–	–
q	Quantity	–	–	–

To give a first idea of how the three phenomena we study relate to the central issue of this paper, consider the English data in Table 2, which includes the data on nominalizations provided earlier in (2).

English has basic interrogative words for most of the semantic categories investigated. An exception is Quantity. This can of course be questioned by means of *how much*, but this is an internally complex expression and not a basic word. In Section 4 we will give our reasons for restricting ourselves to single unanalyzable words. English likewise has to use complex expressions for deictic reference to Quantities (*that much*) and Reasons (*therefore, for that reason*), while it has basic demonstrative words for the other semantic categories. And finally, there are nominalization processes that produce lexical items designating Individuals and Locations, but none that produce lexical items designating Times, Manners, Quantities, and Reasons. This brief illustration shows that even within a single language grammatical processes may show different degrees of sensitivity to semantic categorization.

3. The sample

The sample used in this study consists of 24 languages, distributed across the major families and subfamilies represented in Brazil as shown in Table 3.

To the extent that the limitations as regards the availability of complete grammars allow this, the sample languages are distributed across language families applying the sampling method of Rijkhoff, Bakker, Hengeveld & Kahrel (1993). This method takes the internal complexity of a family tree into account when deciding how many languages should be taken from each family and subfamily. This is done by calculating a diversity value ('DV' in Table 3) on the basis of the structure of the family tree, which then translates into the assignment of a certain number of languages ('No of lgs' in Table 3) to a language family given a certain sample size.

Table 3. The sample

Language Family	DV	No of lgs	Sample language
Arauan	00.00	1	Jarawara
Arawakan	05.33	2	
Northern-Maipuran		1	Tariana
Southern-Maipuran		1	Apurinã
Arutani-Sape	00.00	1	–
Carib	06.25	2	
Northern		1	Apalaí
Southern		1	Hixkaryana
Chapakura-Wanyam	02.00	1	Warí
Creole	02.00	1	Karipuna Creole French
Kanoê	00.00	1	Kanoê
Katukinan	00.00	1	–
Kwaza	00.00	1	Kwaza
Macro-Ge	16.25	5	
Bororo		1	Bororo
Ge-Kaingang		1	Xavante
Rikbaktsa		1	Rikbaktsa
Maku	00.00	1	Hup
Mataco-Guaykuru	00.00	1	Kadiwéu
Mura-Pirahã	00.00	1	Pirahã
Nambikwaran	00.00	1	Sabanê
Pankararú	00.00	1	–
Panoan	07.50	3	
Northern		1	Matís
South Central		1	Yaminawa
Southeastern		1	Shanenawa
Tikuna	00.00	1	–
Trumai	00.00	1	–
Tucanoan	05.00	2	
Central Tucanoan		1	Cubeo
Tupí	13.00	3	
Mundurukú		1	Mundurukú
Tupí-Guaraní		1	Asuriní
Tuxá	00.00	1	–
Witotoan	00.00	1	Bora
Yanomam	00.00	1	Sanumá

Table 3 shows that there is not a perfect match due to various gaps in the sample, but that the distribution across families is non-overlapping so that the sample represents strong internal diversity.

4. Basic question words

The first group of items that may shed light on the distribution of semantic categories across languages is that of basic question words, i.e. unanalyzable words that represent the questioned element in a content question.

In the tradition of Berlin & Kay (1969), we only consider basic terms and not compositional expressions. In the languages studied, both types can be found. Thus in (6) we find a single unanalyzable interrogative word, while in (7) the interrogative expression is derived:

Karipuna Creole French

- (6) *Kōbyā u gā?*
 how.many 2.SG have
 ‘How many do you have?’ (Tobler 1983:42)

Karipuna Creole French

- (7) *ki tā u ke nathe-l?*
 what time 2.SG FUT plait-3.SG
 ‘What time will you plait it?’ (Tobler 1983:43)

In (6) we can observe that *kōbyā* is a basic word, because it is not (synchronically) derived from any other expression and is not combined with any other element. In example (7), on the other hand, the expression *ki + tā* is clearly compositional. Only monomorphemic expressions such as the one illustrated in (6) will be considered. Note that in several of the languages under investigation interrogative and indefinite words are not distinguished from one another (see Hengeveld *et al.* 2007). In these cases we have classified the basic words among these as basic interrogative words.

Basic question words for a variety of semantic categories are illustrated in the following examples:

Apurinã: individual (x)¹

- (8) *Ki-pa apo-pe?*
 who/what-INT arrive-PFV
 ‘Who/What has arrived?’ (Facundes 2000: 362)

Bororo: location (l)

- (9) *Kai-ba?*

where-INT

‘Where?’ (Crowell 1979: 81)

Mundurukú: time (t)

- (10) *Puje o’-ju?*

when 3.SG-go?

‘When did he/she go?’ (Crofts 2004: 232)

Apalaí: manner (m)

- (11) *Otãto t-yri-se ey-a?*

how INF-do-COMPL 3-by

‘How did he do it?’ (Koehn & Koehn 1986: 58)

Sabanê: reason (r)

- (12) *Kate-ay-i-say-al-a.*

why-go-vs-PROG-PRES.NEUT-INT

‘Why is s/he leaving?’ (Araújo 2004: 209)

Karipuna Creole French: quantity (q)

- (13) *Kõbyã u gã?*

how.many 2.SG have

‘How many do you have?’ (Tobler 1983: 42):

In some languages basic question words cover more than one semantic category. This is illustrated for Asuriní in (14) and (15):

Asuriní: manner (m)

- (14) *Mara pa eresan.*

how INT come

‘How did you come?’ (Nicholson 1984: 32)

Asuriní: quantity (q)

- (15) *Mara pa hepya.*

how INT cost

‘How much did it cost?’ (Nicholson 1984: 16)

The overall classification of the availability and distribution of basic question words is given in Table 4.

Table 4 shows a number of very strong implicational hierarchies, which are listed in (16)–(18):

- (16) x ⊂ l ⊂ m ⊂ t

- (17) x ⊂ l ⊂ m ⊂ r

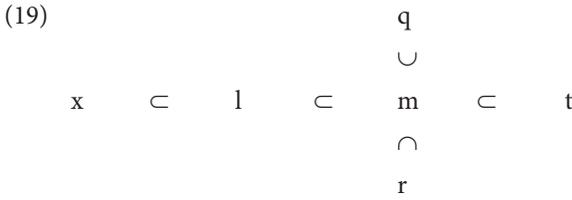
- (18) x ⊂ l ⊂ m ⊂ q

Table 4. Basic question words in the languages of the sample

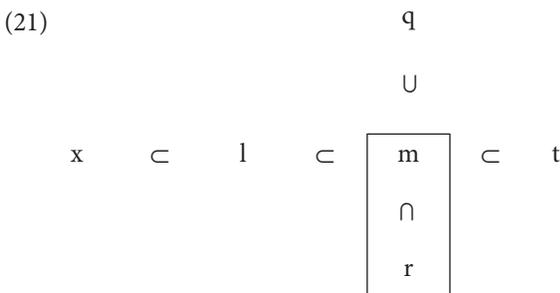
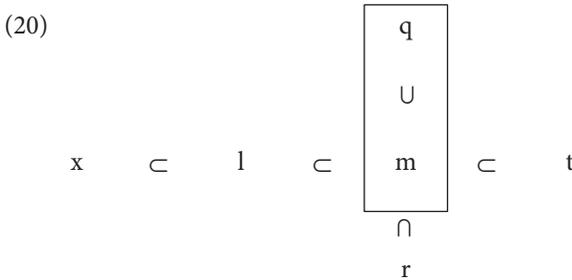
Language	x	l	m	q	r	t
Warí						
Bora	x					
Bororo	x					
Kadiwéu	x					
Kwaza	x					
Matís	x					
Pirahã	x					
Sanumá	x					
Hup	x	l				
Kanoê	x	l				
Tariana	x	l				
Mundurukú	x	l	m			
Apurinã	x	l	m			
Cubeo	x	l	m			
Apalaí	x	l	m	q		
Asuriní	x	l	m/q	m/q		
Hixkaryana	x	l	m	q		
Jarawara	x	l, l/m/q	l/m/q	l/m/q		
Karipuna Creole French	x	l	m	q		
Xavante	x	l	m/q	m/q		t
Rikbaktsa	x	l	r/m		r/m	
Shanenawa	x	l	m		r	
Yaminawa	x	l	r/m		r/m	
Sabanê	x	l	m		r	t

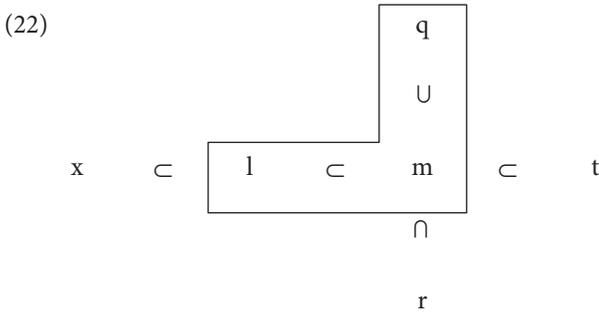
The first three semantic categories are the same: in no language is there a basic question word for manners without there being basic question words for locations and individuals, and in no language is there a basic question word for locations without there being a basic question word for individuals. Furthermore, the presence of the last category of these three, manner, is a precondition for there to be a basic question word for time, reason, or quantity, or a combination of those. However, there are no implicational relationships among the last three categories.

These observations may be combined in the following representation:



Note that one of the possibilities that (19) predicts is that there may be languages with no basic question words at all, and in the sample this is true for Wari. In this language a special construction type is used in which a gap indicates which constituent is being questioned. It is furthermore worth noting that not a single language in the sample has basic question words for all categories. The representation in (19) is inspired by the way semantic maps are organized (see e.g., Haspelmath 2003), though what we have in (19) is an implicational map rather than a semantic map (see Hengeveld & van Lier 2010). Confirmation for the validity of this map may furthermore be found in the fact that in those cases in which languages use a single basic question word for more than one semantic category, these categories are contiguous within the map, as shown in (20)–(22) for the three combinations found.





These results may now be compared to the ones obtained by Mackenzie (2009), a typological study of the distribution of interrogative words based on a world-wide sample of 50 languages. He arrives at the following (statistical) implicational hierarchy:

$$(23) \quad x \subset l \subset q \subset m \subset t \subset r$$

In comparing this hierarchy with the ones in (16)–(18) an important difference shows up. In Mackenzie's the semantic category of quantity is very high in the hierarchy, while in the Brazilian area it occupies a peripheral position.

Mackenzie (2009) furthermore investigated the phonemic weight of basic interrogatives, and ranks the various categories from lower to higher phonemic weight with the result in (24):

$$(24) \quad x - l - m - q - r - t$$

What is remarkable here in comparison with the hierarchies in (16)–(18) is the low position of the semantic category of time as compared to quantity and reason.

These two differences between the Brazilian data and those obtained from a worldwide sample point at an areal aspect to the grammaticalization of semantic categories, an issue to which we will return below.

5. Basic demonstrative words

By basic demonstrative words we mean unanalyzable words that can be used to deictically indicate a semantic category. Using the same criteria for basicness as in the case of basic interrogative words, basic demonstrative words are found for the categories Individual (x), Location (l), Manner (m), and Time (t) in the languages of the sample. Basic demonstrative words for reason and quantity were not encountered, the alternative being composite expressions of the type *for that reason* and *that much*. The following examples illustrate the four types encountered:

- Hup: individual (x)
- (25) *Núp nĩ mɔmbɔʔk ni-ʔeʔ-yǎh-ǎh.*
 this 1.SG.POSS pot be-PERF-FRUST-DECL
 ‘This used to be my pot.’ (Epps 2005: 299)
- Shanenawa: location (x)
- (26) *Jumaj nĩnu-ki.*
 jaguar here-DECL
 ‘The jaguar is here.’ (Cândido 2004: 123)
- Hixkaryana: manner (m)
- (27) *Oske kanye mok nexeye akakheno.*
 thus one.who.says.it that.one he.was my.dead.father
 ‘My dead father was the one who told it like that.’ (Derbyshire 1979: 132)
- Karipuna Creole French: time (t)
- (28) *Anu mǎje ató.*
 HORT eat now
 ‘Let’s eat now!’ (Tobler 1983: 80)

Table 5 shows the distribution of basic demonstrative words across the languages of the sample.

A clear pattern emerges here, that shows that in all but one of the languages of the sample the following hierarchy holds:

- (29) $x \subset l \subset m \subset t$

This is of course the hierarchy that we presented in (16) in relation to interrogative words.

The peripheral status of time (t) in this hierarchy is reflected in the fact that the only instances found for temporal demonstratives are the translational equivalents of ‘now’, never of ‘then’, which is rather realized as ‘at that time’.

There are quite a number of languages, given at the top of Table 5, that have no demonstratives at all for the categories investigated. These are languages that express demonstrativity in separate elements, that can only be used in combination with a lexical element, as in the following example:

- Warí
- (30) *carawara ma*
 animal DEM:M/F:DISTAL
 ‘that animal’ (Everett & Kern 1997: 153)

It is the combination of the head noun *carawara* with the demonstrative that designates an individual here. The demonstrative element could not be used on its

Table 5. Basic demonstrative words in the languages of the sample

Language	x	l	m	t
Kwaza				
Mundurukú				
Sabanê				
Warí				
Yaminawa				
Cubeo	x			
Hup	x			
Kanoê	x			
Pirahã	x			
Sanumá	x			
Apalaí	x	l		
Apurinã	x	l		
Jarawara	x	l		
Rikbaktsa	x	l		
Shanenawa	x	l		
Hixkaryana	x	l	m	
Tariana	x	l	m	t
Bororo	x	l	m	t
Bora	x	l	m	t
Karipuna Creole French	x	l	m	t
Xavante	x	l	m	t
Matís		l		
Asuriní	?	?	?	?
Kadiwéu	?	?	?	?

own with that same designation. The language thus has demonstrative elements, but not ones that correspond to any of the semantic categories investigated.

A special case in Table 5 is Matís. This language has basic locative demonstrative words designating locations, as shown in the following examples:

- Matís
 (31) *Şapu-Ø ni tsad-e-k*
 Şapu-ABS here sit-NONPAST-DECL
 ‘Şapu is sitting here.’ (Ferreira 2005:152)

Matis

- (32) *Tumi-Ø u duke-e-k*
 Tumi-ABS there lie-NONPAST-DECL
 ‘Tumi is lying there.’

The same demonstrative bases can be found in demonstratives denoting individuals such as *nikid* ‘this one’ and *ukid* ‘that one’. Since *-kid* is a productive nominalizer in the language (Ferreira 2005: 83), it seems reasonable to analyse these elements as internally complex, meaning something like ‘the one here’ and ‘the one there’. If this analysis is correct, Matis is a counterexample to the hierarchy in (29). It is interesting to note that in the closely related language Matses the corresponding demonstratives are also historically nominalized forms, but in this language the nominalizer is no longer productive (Fleck 2003: 259), which means that the resulting forms should be analyzed as basic demonstratives referring to individuals.

A final observation that deserves mention and that follows from a comparison of Tables 4 and 5 is that for every language, again with the exception of Matis, the set of semantic categories covered by basic demonstrative words is always a subset of or identical to the set of semantic categories covered by basic interrogative words.

These results may now be compared to the ones obtained by George (2010), a typological study of the distribution of demonstrative words based on a world-wide sample of 50 languages. She arrives at the following (statistical) implicational hierarchy:

- (33) $x \subset l \subset t \subset m \subset r \subset q$

In comparing this hierarchy with the one in (29) two differences show up. In George’s sample (i) there are languages which do have basic demonstrative words designating reasons and quantities, something which we did not find in the Brazilian area; (ii) the position of basic demonstrative words designating times and those designating manners are inverted. The latter fact seems to indicate that the temporal category has a more peripheral status in Brazilian languages than in the languages of George’s sample.

George (2010) furthermore investigated the phonemic weight of basic demonstratives, and ranks the various categories from lower to higher phonemic weight with the result in (34):

- (34) $x - l - m - t - r - q$

Interestingly, in this case the four initial categories in the scale occupy the same order that these categories occupy in our hierarchy in (29).

In all, the comparison of the Brazilian data with those obtained for a world-wide sample again point at areal influences involving, as in the case of interrogative words, the categories of time and quantity.

6. Nominalization

A third aspect of grammar that may reveal variation between languages as regards semantic categorization is the process of nominalization, which is the process by which nouns are derived from another class of lexical elements. The derived nouns may designate various types of semantic categories. Within our sample, and for the semantic categories investigated, nominalization produces nouns designating individuals (x), locations (l), and times (t). Virtually all languages in the sample also have a nominalization process that produces nouns designating States-of-Affairs (e). As explained in the introduction, we have excluded these from the investigation, as similar information was impossible to obtain for basic interrogative and demonstrative words.

The following examples illustrate the three types of nominalization:

- Kwaza: individual (x)
- (35) a. *'wi-ki*
cut-DECL
'He cuts.'
- b. *wi-ni'te*
to cut-INSTR.NR
'knife' (Van der Voort 2004: 483):
- Asuriní: location (l)
- (36) a. *osoka*
kill
'to kill'
- b. *ise i-sokat-awa*
I.SG IND-kill-LOC.NR
'the place where I killed him' (Nicholson 1984: 49)
- Shanenawa time (t)
- (37) a. *şuşu*
play
'to play'
- b. *şuşu-tian*
play-TEMP.NR
'break, playtime' (Cândido 2004: 78)

Various languages have nominalization strategies that may produce nouns designating more than one type of semantic category. Such a language is Yaminawa, as shown by the following examples:

- Yaminawa: individual (x) and location (l)
- (38) a. *mâtso*
sweep
'to sweep'
- b. *mâtso-ti*
sweep-NR
'broom'

- (39) a. *nashi*
 bath
 'to bath'
- b. *nashi-ti*
 bath-NR
 'shower' (Faust 2002: 96):

Table 6 shows the distribution of the processes of nominalization with respect to the various semantic categories across the languages of the sample.

The hierarchy that emerges from Table 6 is given in (40):

- (40) $x \subset l \subset t$

This hierarchy explains the distribution, and again, in cases in which the same strategy is used for different semantic categories, the overlap concerns contiguous

Table 6. Nominalizations in the languages of the sample

Language	x	l	t
Karipuna Creole French			
Warí			
Bororo			
Sanuma			
Cubeo	x		
Hup	x		
Kadiweu	x		
Kanoê	x		
Kwaza	x		
Munduruku	x		
Pirahã	x		
Sabanê	x		
Apurinã	x	l	
Jarawara	x	l	
Matís	x	l	
Asurini	x	l	
Tariana	x	l	
Xavante	x	l	
Yaminawa	x/l	x/l	
Apalaí	x	l/t	t
Hixkaryana	x	l	t
Shanenawa	x/l	x/l	t
Bora	?	?	?
Rikbaktsa	?	?	?

categories on the hierarchy. But a category prominent in the previous hierarchies is missing: in the earlier hierarchies manner (m) occupies a position in between location (l) and time (t).

There seems to be a straightforward explanation for this fact. Manner is first and foremost a semantic category that is used to modify. In sentences like *How did he do it* and *He did it thus* the manner expression modifies the main verb, as is evident from lexical equivalents such as *He did it silently*. The process of nominalization produces nouns, which are primarily used as heads of phrases, not as modifiers.

7. Conclusion

In this paper we have looked at three different phenomena that are sensitive to semantic categorization. Taken together, the extent to which the grammars of the languages studied are sensitive to the semantic categories studied can be captured in the hierarchy given in (19) and repeated here as (41):

$$(41) \quad \begin{array}{ccccccc} & & & & q & & \\ & & & & \cup & & \\ x & \subset & l & \subset & m & \subset & t \\ & & & & \cap & & \\ & & & & r & & \end{array}$$

The implicational map in (41) as a whole is needed to describe the distribution of basic interrogative words. With one exception, its horizontal axis appropriately describes the distribution of basic demonstrative words, which were furthermore shown to always designate a subset of the semantic categories designated by basic interrogative words within the same language. Finally, nominalizations were shown to display a somewhat different behaviour, in the sense that the horizontal implicational hierarchy seems to be relevant to them, but that manner (m) does not show up as a relevant category. In all, we feel that the results of this typological investigation confirm the relevance of semantic categorization in grammar, as explicitly captured in the FDG model, and that the differences between languages as regards the extent to which these categories play a role in the organization of the grammatical system are far from random.

We have noted in the comparison of our results with the ones obtained for interrogative words in Mackenzie (2009) and George (2010) that there are a number of differences that indeed seem to point at an areal influence in semantic categorization. The two categories that showed this influence most clearly are the category

of quantity, which ranks lower in the Brazilian area than it does in a world-wide sample for basic interrogative words, and the category of time, which ranks lower in the Brazilian area for basic demonstrative words.

Received 23 August 2009.

Revised version 21 September 2010.

Notes

* We are grateful to J. Lachlan Mackenzie, Hein van der Voort and the editors of *Functions of Language* for their comments on an earlier version of this paper.

1. In this and the following example the interrogative suffix is not part of the question word but a general interrogative marker.
2. The criteria used in Mackenzie (2009) for establishing the basic status are slightly different from ours. We have reclassified Mackenzie's data in terms of our criteria, with only marginal differences in the outcome. We are grateful to Lachlan Mackenzie for his help in analyzing the data.

References

- Araujo, Gabriel Antunes de. 2004. *A grammar of Sabanê: A Nambikwaran language*. Utrecht: LOT.
- Berlin, Brent & Paul Kay. 1969. *Basic color terms*. Berkeley, CA: University of California Press.
- Cândido, Gláucia Vieira. 2004. *Descrição morfossintática da língua Shanenawa (Pano)*. Campinas, SP: UNICAMP dissertation.
- Carlson, Robert. 1994. *A grammar of Supyire* (Mouton Grammar Library 14). Berlin: Mouton.
- Crofts, Marjorie. 2004. *Aspectos da língua Munduruku*. Cuiabá: Summer Institute of Linguistics.
- Crowell, Thomas H. 1979. *A grammar of Bororo*. Ithaca, NY: Cornell University dissertation.
- Derbyshire, Desmond C. 1979. *Hixkaryana*. *Lingua Descriptive Studies*, 1. Amsterdam: North-Holland.
- Epps, Patience. 2005. *A grammar of Hup*. Berlin: Mouton.
- Everett, Daniel & Barbara Kern. 1997. *Warí: The Pacaás Novos language of Western Brazil*. London: Routledge.
- Facundes, Sidney da Silva. 2000. *The language of the Apurinã people of Brazil (Arawak)*. Buffalo, NY: SUNY-Buffalo dissertation.
- Faust, Norma & Eugene E. Loos. 2002. *Gramática del idioma Yaminahua*. Lima: Instituto Lingüístico de Verano.
- Ferreira, Rogerio Vicente. 2005. *Lingua Matis (Pano): Uma descrição gramatical*. Campinas, SP: UNICAMP dissertation.
- Fleck, David William. 2003. *A grammar of Matses*. Houston, TX: Rice University dissertation.
- George, Jessica. 2010. *Demonstratives and semantic categorization*. Amsterdam: University of Amsterdam MA thesis.

- Haspelmath, Martin. 2003. The geometry of grammatical meaning: Semantic maps and cross-linguistic comparison. In Michael Tomasello (ed.), *The new psychology of language* 2. 211–243. New York, NY: Erlbaum.
- Hengeveld, Kees & Niels Smit. 2009. Dynamic formulation in Functional Discourse Grammar. In Kees Hengeveld & Gerry Wanders (eds.), *Semantic representation in Functional Discourse Grammar*. *Lingua* 119(8). 1118–1130.
- Hengeveld, Kees & Eva van Lier. 2010. The implicational map of parts-of-speech. In Andrej Malchukov, Michael Cysouw & Martin Haspelmath (eds.), *Semantic maps: Methods and applications*. *Linguistic Discovery* 8(1). 129–156.
- Hengeveld, Kees & J. Lachlan Mackenzie. 2008. *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford: OUP.
- Hengeveld, Kees, Eli Nazareth Bechara, Roberto Gomes Camacho, Alessandra Regina Guerra, Taisa Peres de Oliveira, Eduardo Penhavel, Erotilde Goreti Pezatti, Liliane Santana, Edson Rosa Francisco de Souza & Maria Luiza Teixeira de Sousa. 2007. Basic illocutions in the native languages of Brazil. In Marize Mattos Dall'Aglio Hattner & Kees Hengeveld (eds.), *Advances in Functional Discourse Grammar*. *Alfa-Revista de Lingüística* 51(2). 73–90.
- Koehn, Edward H. & Sally S. Koehn. 1986. Apalai. In Desmond C. Derbyshire & Geoffrey K. Pullum (eds.), *Handbook of Amazonian languages* 1. 33–127. Berlin: Mouton.
- Mackenzie, J. Lachlan. 2009. Content interrogatives in a sample of 50 languages. In Kees Hengeveld & Gerry Wanders (eds.), *Semantic representation in Functional Discourse Grammar*. *Lingua* 119(8). 1131–1163.
- Nicholson, Velda 1984. *Aspectos da língua Asuriní*. Cuiabá: Summer Institute of Linguistics.
- Rijkhoff, Jan, Dik Bakker, Kees Hengeveld & Peter Kahrel. 1993. A method of language sampling. *Studies in Language* 17(1). 169–203.
- Tobler, S. Joy. 1983. *The grammar of Karipuna Creole*. Série Lingüística, 10. Brasília: Summer Institute of Linguistics.
- Voort, Hein van der. 2004. *A grammar of Kwaza* (Mouton Grammar Library 29). Berlin: Mouton.

Authors' addresses

Kees Hengeveld (corresponding author)
 Department of Theoretical Linguistics
 Universiteit van Amsterdam
 Spuistraat 210
 1012 VT Amsterdam, The Netherlands
 p.c.hengeveld@uva.nl

Maria Luiza Braga, Elisiene de Melo Barbosa, Jaqueline Silveira Coriolano, Juliana Jezuiño da Costa, Mariana de Souza Martins, Diego Leite de Oliveira, Vinícius Maciel de Oliveira, Luana Gomes Pereira, Cassiano Luiz do Carmo Rodrigues, Viviane dos Ramos Soares
 Universidade Federal do Rio de Janeiro
 Departamento de Linguística e Filologia
 Av. Horácio Macedo 2151
 21941-917 Rio de Janeiro RJ, Brazil

malubraga@terra.com.br

jaquelinecor@yahoo.com.br

Mariana_martins1@yahoo.com.br

vmoliveira@vista.aero

cassianolsantos@ibest.com

elisiene@hotmail.com

julianajezuino@yahoo.com.br

diegoleitede@gmail.com

luaninhagomes@gmail.com

viviane_soares@yahoo.com.br

Liliane Santana

Universidade Estadual Paulista — Campus São José do Rio Preto IBILCE

R. Cristóvão Colombo 2265

Jardim Nazarte

15054-000 São José do Rio Preto SP, Brazil

lilianasantana@uol.com.br