Semantics

Cornelia Endriss[1]*, Stefan Hinterwimmer*[2]*, Stavros Skopeteas*[1]
University of Potsdam ([1]) and Humboldt University Berlin ([2])

The guidelines for semantics comprise a number of layers related to quantificational structures as well as some crucial semantic properties of NPs with respect to information structure: definiteness, countability, and animacy.

1 Preliminaries

Those features that are decisive for the semantic interpretation of a sentence have to be represented. We assume that syntactic annotation has already taken place so that all relevant syntactic features which are also interesting for the semantic level are explicit already.

The present guidelines were developed for annotating elements that occur in a corpus text. Elements that do not form part of the archived data, but arise from the analysis of the data (as in the case of ellipsis, traces, etc.) are not supported in the current version.

2 Layer Declaration

Table 1: Layers

Layer	Name
Quantificational properties	QuP
Interpretation of adverbially quantified structures	IN_adv

Interdisciplinary Studies on Information Structure 07 (2007): 135–145
Dipper, S., M. Götze, and S. Skopeteas (eds.):
Information Structure in Cross-Linguistic Corpora

Interpretation of possibly ambiguous quantified structures	IN_quant
Definiteness properties	DefP
Countability	C
Animacy	A

Table 2: Tagset declaration

Layer	Tags	Short Description
Quantificational properties (QuP)	ALL	universal quantifier
	EXIST	existential quantifier
	GEN	generic quantifier
	NUM	numerals
	Q	other quantifier
Interpretation of adverbially quantified structures (IN_adv)	N	nucleus
	QADV	quantificational adverbial
	R	restrictor
Interpretation of possibly ambiguous quantified structures (IN_scope)	ALL	universal quantifier
	EXIST	existential quantifier
Definiteness properties (DefP)	GEN	generic
	SP	specific
	U	unique
	USP	unspecific
Countability (C)	C	count
	M	mass
Animacy (A)	A	animate, non-human
	H	animate, human
	I	inanimate
	IA	unclear if animate or inanimate

3 Layer I: Quantificational properties (QuP)

3.1 Preliminaries

This layer deals with the annotation of quantificational elements and the resulting interpretation of those parts or the sentences containing those parts.

3.2 Tagset declaration

The semantic annotation has to enable queries combining quantificational and syntactic properties, e.g. “search for quantificational DPs”; “search for existential adverbs”, etc.

Assuming that syntactic information is provided by the layers that describe constituent structure, the semantic annotation contains the following labels:

Table 2: Tags for quantificational properties

tag	description	markable
ALL	universal quantifier	quantificational NPs/adverbials
EXIST	existential quantifier	quantificational NPs/adverbials
GEN	generic quantifier	covert operator
Q	other quantifiers	quantificational NPs/adverbials

3.3 Illustrative examples

If we assume an annotation layer ‘CS’ that displays constituent structures:

(1) English

<WORDS>	every	girl	likes	some	horse
<CS>	NP			NP	
<QuP>	ALL			EXIST	

(2) English

<WORDS>	dogs	always	have	green	eyes
<CS>	NP	ADV	VP		
<QuP>		ALL			

(3) English

<WORDS>	no	one	saw	three	horses
<CS>	NP			NP	
<QuP>	Q			Q	

A covert generic operator should be annotated whenever a sentence gets a generic interpretation. This can be tested in following way: Whenever *always/generally* can be inserted without changing the intended interpretation, a generic covert operator can be assumed:

(4) English

<WORDS>	a	dog		has	a	tail
<CS>	NP		ADV	VP		
<QuP>			GEN			

(5) English

<WORDS>	dogs		have	tails
<CS>	NP	ADV		
<QuP>		GEN		

This should not be confused with the existential interpretation that bare plurals often get:

(6) English

<WORDS>	dogs	were	sleeping
<CS>	NP	VP	
<QuP>	EXIST		

4 Layer II: Interpretation of adverbially quantified structures (IN_ADV)

4.1 Preliminaries

This layer deals with the annotation of the relation of restrictor and nucleus in sentences with quantificational adverbials.

4.2 Tagset declaration

Table 3: Tags for interpretation adverbially quantified structures

tag	description	markable
N	nucleus	part of sentences with Q-Adverbs
QADV	quantificational adv	adverbial
R	restrictor	part of sentences with Q-Adverbs

4.3 Illustrative example: Adverbially quantified structures interpretation

(7) English

<WORDS>	dogs	always	have	green	eyes
<IN_adv>	R	QADV	N		

5 Layer III: Interpretation of possibly ambiguous quantified structures (IN_scope)

5.1 Preliminaries

This layer deals with the annotation of the interpretation of quantificational elements, i.e. the scope of DP-quantifiers.

5.2 Tagset declaration

The units to be annotated are possibly ambiguous sentences that contain quantificational elements. The possible reading(s) of these sentences should be marked in the <IN_scope> field.

Table 4: Tags for interpretation of possibly ambiguous quantified structures

tag	description	markable
ALL	universal quantifier	quantificational NPs/adverbials
EXIST	existential quantifier	quantificational NPs/adverbials
GEN	generic quantifier	covert operator
Q	other quantifiers	quantificational NPs/adverbials
>	has scope over	sentences

5.3 Illustrative example: Scope interpretation

(8) English

<WORDS>	every	girl	likes	some	horse
<IN_scope>	ALL> EXIST; EXIST>ALL				

6 Layer IV: Definiteness properties (DefP)

6.1 Preliminaries

This layer contains information about definiteness. Definiteness encoded (e.g. through articles) is given in the morphemic translation (of the article).

6.2 Tagset Declaration

Table 5: Tags for definiteness properties

tag	description	markable
GEN	generic	NP (Indefinites/Definites)
SP	specific	NP (Indefinites)
U	unique	NP (Definites)
USP	unspecific	NP (Indefinites)

6.3 Instructions and illustrative examples

Annotate as definite:

- definite articles: *the*
- demonstratives: *this*
- possessives: *your horse*, *his book*

Annotate as indefinite:

- indefinite articles: *a*

(9) English

<WORDS>	Peter	is	looking	for	some	horse
<DefP>						USP

Test for unspecificity: The respective sentence could e.g. be followed by *And it does not matter which one.*

(10) English

<WORDS>	Peter	is	looking	for	some	horse
<DefP>					SP	

Test for specificity: The respective sentence could e.g. be followed by *But he has not found it yet.*

Kind Interpretations (Note the difference to generic interpretations that should be annotated as originating from a covert operator, cf. Section 2):

(11) English

<WORDS>	der Dinosaurier / Dinosaurier	ist / sind	ausgestorben
<DefP>	GEN		

The markables are DPs (and not single definite/indefinite markers): if more definite and indefinite markers occur in the same DP, only the resulting definiteness is annotated:

(12) English

<WORDS>	the	three	cowboys
<DefP>	U		

DPs can, of course, be stacked and should be annotated as such. (In Exmeralda, this can only be done by supplying for multiple DefP layers):

(13) English

<WORDS>	the	mother	of	the	boys
<DefP1>	U			U	
<DefP2>	U				

The respective DP should only be annotated as unique if the text allows us to conclude that the object denoted by the DP is the only object for which the property described by the corresponding NP holds. (In the literature it is sometimes claimed that non-unique definites exist.)

7 Layer V: Countability (C)

At this layer we encode information concerning the entity type (count/mass). The markables for this information are nouns. Nouns/DPs that turn up as part of sayings do not have to be annotated. (Looking at German, there are many

sayings or phrases that combine with bare singulars such as “in Frage stellen”, which regularly do not exist in German. These should not be annotated.)

7.1 Tagset Declaration

The following abbreviations are used for the annotation of the count/mass property of the noun/NP.

Table 6: Tags for countability

tag	description	markable
C	count	noun/DP
M	mass	noun/DP

7.2 Instructions and illustrative examples

(14) English

<WORDS>	cat
<C>	C

(15) English

<WORDS>	milk
<C>	M

8 Layer VI: Animacy (A)

8.1 Preliminaries

At this layer we encode information concerning the animacy. Since this annotation layer will be especially interesting for corpus studies concerning the impact of animacy on word order, topicality, and related issues, we adopt a rather detailed classification, so that users of the database are able to specify in

their queries which kind of NPs they want to count as animates or inanimates. The markables for animacy are nouns (both proper or common nouns).

8.2 Tagset Declaration

The following abbreviations are used for the annotation of the animacy property of the noun.

Table 7: Tags for animacy

tag	description	markable
A	animate, non-human	noun
H	animate, human	noun
I	inanimate	noun
IA	inanimate/animate	noun

8.3 Instructions and illustrative examples

Clear instances of human beings are annotated as ‘h’:

(16) English

<WORDS>	woman
<A>	H

Clear instances of non-human animates are annotated as ‘a’:

(17) English

<WORDS>	cat
<A>	A

Clear instances of inanimates are annotated as ‘i’:

(18) English

<WORDS>	milk
<A>	I

The following categories concern types of entities that are not clear instances of the above categories. Since it depends on the criteria of a certain study whether each of these categories should be treated as animate or inanimate or if it should simply be excluded from the query, we recommend grouping these in the remaining category IA:

- body parts:

(19) English

<WORDS>	hand
<A>	IA

- non-humans with human-like properties. These referents are not humans, but they may have similar properties to humans in several respects (agenthood, shape, motion) and may be treated like humans in certain languages.

(20) English

<WORDS>	robot
<A>	IA