

# The *TOSCA-ICLE* Tagset

## *Tagging Manual*

*Jan Aarts, Henk Barkema, and Nelleke Oostdijk*

This manual accompanies the TOSCA-ICLE tagger/lemmatizer version 1.0

TOSCA Research Group for Corpus Linguistics  
Dept. of Language and Speech  
Faculty of Arts  
University of Nijmegen  
The Netherlands

e-mail: [tosca@let.kun.nl](mailto:tosca@let.kun.nl)  
www: <http://lands.let.kun.nl/TSpublish/tosca/>

Version 1.0 - April 1997

© TOSCA Research Group. This document is subject to copyright. No part of it may be copied without the express, written permission of the TOSCA Research Group



## CONTENTS

<b>1. INTRODUCTION.....</b>	<b>5</b>
1.1 THE TOSCA-ICLE TAGGER/LEMMATIZER.....	5
1.2 THE TOSCA-ICLE TAG SELECTION TOOL.....	6
1.3 THE TOSCA-ICLE TAGSET .....	6
<b>2. MAJOR WORDCLASSES, THEIR SUBCLASSES AND FEATURES .....</b>	<b>9</b>
2.1 ADJECTIVE .....	9
2.1.1 <i>General adjective</i> .....	9
2.1.2 <i>Ordinal adjective</i> .....	12
2.2 ADVERB .....	13
2.2.1 <i>General adverb</i> .....	13
2.2.2 <i>Connective adverb</i> .....	15
2.2.3 <i>Negative adverb</i> .....	16
2.2.4 <i>Phrasal adverb</i> .....	17
2.2.5 <i>Wh-adverb</i> .....	17
2.3 ARTICLE.....	18
2.4 CONJUNCTION .....	18
2.4.1 <i>Co-ordinating conjunction</i> .....	18
2.4.2 <i>Subordinating conjunction</i> .....	19
2.5 EXISTENTIAL <u>THERE</u> .....	21
2.6 NOUN.....	21
2.6.1 <i>Collective nouns</i> .....	28
2.7 NOMINAL ADJECTIVE.....	29
2.8 NUMERAL .....	30
2.8.1 <i>Cardinal numeral</i> .....	31
2.8.2 <i>Ordinal numeral</i> .....	31
2.8.3 <i>Fractional numeral</i> .....	32
2.8.4 <i>Hyphenated numeral</i> .....	32
2.8.5 <i>Multiplicative</i> .....	32
2.9 PREPOSITION .....	33
2.9.1 <i>General preposition</i> .....	33
2.9.2 <i>Phrasal preposition</i> .....	36
2.10 PROFORM .....	37
2.10.1 <i>Pro-conjoin</i> .....	37
2.10.2 <i>Proform one</i> .....	37
2.10.3 <i>Proform so</i> .....	38
2.11 PRONOUN .....	38
2.11.1 <i>Anticipatory it</i> .....	39
2.11.2 <i>Assertive pronoun</i> .....	40
2.11.3 <i>Cleft it</i> .....	40
2.11.4 <i>Demonstrative pronoun</i> .....	41
2.11.5 <i>Exclamatory pronoun what</i> .....	41
2.11.6 <i>Interrogative pronoun</i> .....	41
2.11.7 <i>Negative pronoun</i> .....	42
2.11.8 <i>Nominal possessive pronoun</i> .....	42
2.11.9 <i>Non-assertive pronoun</i> .....	43
2.11.10 <i>Pronoun one</i> .....	43
2.11.11 <i>Personal pronoun</i> .....	43
2.11.12 <i>Possessive pronoun</i> .....	44
2.11.13 <i>Quantifying pronoun</i> .....	45
2.11.14 <i>Reciprocal pronoun</i> .....	45

2.11.15 Relative pronoun.....	45
2.11.16 -self pronoun.....	46
2.11.17 Pronoun such .....	46
2.11.18 Universal pronoun.....	46
2.12 PARTICLE .....	47
2.12.1 Particle for.....	47
2.12.2 Particle to.....	47
2.12.3 Particle with.....	48
2.13 VERB.....	48
2.13.1 Auxiliary verb .....	49
2.13.1.1 Do: auxiliary of periphrasis and emphasis .....	50
2.13.1.2 Passive auxiliary .....	52
2.13.1.3 Progressive auxiliary.....	53
2.13.1.4 Perfective auxiliary .....	54
2.13.1.5 Modal auxiliary.....	55
2.13.1.6 Semi-auxiliary.....	56
2.13.1.7 Semi-auxiliary followed by -ing participle .....	57
2.13.2 Lexical verb .....	59
2.13.2.1 Intransitive verb .....	61
2.13.2.2 Copular verb .....	62
2.13.2.3 Monotransitive verb .....	62
2.13.2.4 Ditransitive verb .....	64
2.13.2.5 Dimono transitive verb.....	64
2.13.2.6 Complex transitive verb .....	65
2.14 MISCELLANEOUS.....	67
2.14.1 Discourse item .....	67
2.14.2 Foreign word or expression.....	69
2.14.3 Interjection .....	69
2.14.4 Prepositioned bound morpheme (prefix).....	70
2.14.5 Postpositioned bound morpheme (suffix).....	70
2.15 GENITIVE MARKER .....	71
2.16 PUNCTUATION .....	71
<b>3. REFERENCES.....</b>	<b>73</b>
<b>APPENDIX 1: ALPHABETICAL LIST OF WORDCLASS AND FEATURE LABELS.....</b>	<b>74</b>
<b>APPENDIX 2: INVENTORY OF TOSCA/ICLE TAGS.....</b>	<b>77</b>

## 1. INTRODUCTION

### Manual

This manual accompanies the TOSCA-ICLE tagger/lemmatizer, which was developed by the TOSCA Research Group for the tagging of the International Corpus of Learner English (ICLE). The manual serves two purposes. The first is to inform the linguist exploring a corpus tagged with the TOSCA-ICLE tagset about the meaning of, and the ideas behind, the tag labels. Its second purpose is to advise linguists in the process of tag selection (see below, section 1.1).

The contents of this manual is as follows: in the remainder of this introduction, we first give a brief description of the tagging process, and then describe the format of the tags, the relation between tokens and tags, between tags, wordclasses and features, the notion of multi-token unit, textual mark-up and the relation between tagging and syntax. The second part of the manual provides a detailed description of the tags in the tagset. Each of the major wordclasses is defined or described and an inventory is provided of the possible combinations of the wordclasses with their features. To enhance clarity, we indicate what the distinctions are between adjacent classes wherever appropriate. Cross-references are to be found in several sections.

### ICLE

The TOSCA-ICLE tagset was designed for the automatic part-of-speech (POS) tagging of subcorpora of the International Corpus of Learner English. The coordinator of the ICLE project is Professor Sylviane Granger of the University of Louvain in Louvain-la-Neuve, Belgium.

### TOSCA

The tagset, tagger/lemmatizer, tag selection tool and tagging manual were produced by the TOSCA Research Group for Corpus Linguistics of the University of Nijmegen, The Netherlands, led by Professor Jan Aarts.

#### ***1.1 The TOSCA-ICLE tagger/lemmatizer***

The tagger/lemmatizer associates with each token in the corpus a tag indicating (minimally) its wordclass membership, while additional feature information may refer to its specific morphological and/or semantico-syntactic characteristics. In addition it assigns a lemma to the token.

#### ***The tagging process***

##### **Tokenization**

Before a text can be tagged with the TOSCA-ICLE tagger, the sentence boundaries must be identified as well as the 'tokens' in them: the word forms and punctuation marks. This activity, which is carried out by a program which is largely rule-based, is called 'tokenization'. The program uses knowledge about punctuation and capitalization conventions, but also relies on statistical information, for example about abbreviations and sentence-initial words.

##### **The lexicon**

TOSCA taggers make use of lexicons in which all possible word forms are listed. Over the years, the word form lexicons for English have been built up. Various sources have contributed to these lexicons, such as tagged corpora and machine-readable dictionaries. This has resulted in extensive lists. The lexicon of the TOSCA-ICLE tagger currently contains about 160,000 token-tag pairs, covering about 90,000 types. Despite this large number of

token-tag combinations the taggers occasionally come across unlisted words. In such instances special components are activated. These try to determine the most appropriate tag on the basis of specific properties of the word, such as its final character string.

### Assignment of most likely tags

After a token with its tags has been called up from the lexicon, the tagger determines which of the tags is most likely in the context in which the token was found. First, information from the lexicon about the probability of a token-tag pair is used to make an initial likelihood ordering of the tags for each token; then this ordering can be adjusted, depending on the probability of a token-tag pair when considering the context of the specific token. This procedure is based on a Hidden Markov model (HMM). Finally, a rule-based component corrects systematic errors made by the statistical components. For further information about the tagger/lemmatizer, please consult its manual (de Haan and van Halteren, 1997).

### 1.2 The TOSCA-ICLE tag selection tool

The tag selection tool can be used to inspect the output of the tagger. In addition, it can be used to select alternatives proposed by the tagger in cases where the tagger has not produced the contextually correct tag as its first option. In a few cases even a contextually appropriate alternative has not been produced by the tagger. In such cases the linguist can make a selection from the entire tagset. For further information about the tag selection tool, please consult its manual.

### 1.3 The TOSCA-ICLE tagset

The TOSCA-ICLE tagset consists of 16 major wordclasses. These major wordclasses may further be specified by features for subclasses as well as for a variety of syntactic, semantic and morphological characteristics. The number of different tags is 219.

The major wordclasses with their subclasses are the following:

Major wordclass	label	subclasses
adjective	ADJ	general, ordinal
adverb	ADV	connective, general, negative, phrasal, <i>wh</i> -
article	ART	definite, indefinite
conjunction	CONJUNC	coordinating, subordinating
existential <i>there</i>	EXTHERE	-
noun	N	collective
nominal adjective	NADJ	-
numeral	NUM	cardinal, fractional, hyphenated, multiplicative, ordinal
preposition	PREP	general, phrasal
proform	PROFM	pro-conjoin, <i>one</i> , <i>so</i>
pronoun	PRON	anticipatory <i>it</i> , assertive, cleft <i>it</i> , demonstrative, exclamatory, interrogative, negative, non-assertive, nominal possessive, <i>one</i> , personal, possessive, quantitative, reciprocal, relative, <i>-self</i> , <i>such</i> , universal
particle	PRTCL	<i>for</i> , <i>to</i> , <i>with</i>
verb	VB	auxiliary verb, lexical verb
miscellaneous	MISC	discourse, foreign word, suffix, prefix
genitive marker	GENM	-
punctuation	PUNC	full stop, exclamation mark, etc.

In addition to the major wordclass-tags listed above, there are two tags for the tagging of extra-textual material:

### Tags for extra-textual material

MARKUP                      for SGML markers  
 UNTAG                        for all tokens not belonging to the text

### Multi-token units

Many combinations of tokens can be regarded as multi-token units. Traditionally they are classified as such when they are lexically, morphologically and syntactically fixed and have institutionalised meanings and/or discourse functions. The philosophy behind the TOSCA-ICLE tagset, however, holds that lexicalisation is a separate level of linguistic description, and that tokens should be tagged for their wordclasses, unless their classification is problematic. This also goes for tokens with slashes, where the slash is regarded as a coordinator. On the other hand, hyphenated wordforms are regarded as one token.

### Criteria for multi-token units

A multi-token unit is a sequence of two or more tokens which forms a unit which can be classified as one of the major wordclasses. A multi-token unit consists of tokens which are difficult to accommodate in separate wordclasses. They are:

1. lexicalised foreign expressions. Examples are *a priori*, *ad hoc*, *ex libris*, *à la carte*, *bona fide*, *de facto*, *de luxe*, *per diem*, *viva voce*, *da capo*, *et al*, *per capita*, *par excellence*.
2. 'absent sense' expressions: expressions with at least one token which has no separate basic sense of its own, which only occurs in one specific lexicalised expression, and which therefore has no separate entry in dictionaries. Examples are: *cos lettuce*, *in lieu of*, *lo and behold*.
3. expressions where the wordclass of one or more of the individual items is uncertain. Examples are: *as well*, *at all*, *of course*.

### Ditto tags

Multi-token units are tagged with 'ditto tags'; each of these is composed of the wordclass (and possibly the features) of the unit, followed by a numerical code which consists of the following:

1. a number for the sequential position of the token in the multi-token unit
2. a number which indicates the number of tokens in the unit

The format is: colon, sequential number, slash, total number.

Example:

as	ADV(connec):1/2	→	first part of the connective adverb
well	ADV(connec):2/2	→	second part of the connective adverb

During tag selection, you should make sure that the numbers are consistent and correct.

### Discontinuity

If a multi-token unit is interrupted by a token from another wordclass, it is discontinuous. In such a case, each token which belongs to the unit is tagged in the same way as when no



interruption had taken place. An example is: *Is she ever going to listen*, which is tagged as follows:

Is	VB(aux,semi,pres,disc):1/3
she	PRON(pers,sing)
ever	ADV(ge,pos)
going	VB(aux,semi,pres,disc):2/3
to	VB(aux,semi,pres,disc):3/3
listen	VB(lex,intr,infin)

In principle all multi-token units can be discontinuous. In our inventory of tags we have not included the features for discontinuity.

### Relation between tags and syntax

The assignment of appropriate tags depends, among other things, on the syntactic contexts of the tokens. For example, the word *after* can be a preposition (e.g. *after the event*), an adverb (e.g. *she arrived the morning after*) or a subordinating conjunction (*After she had left, we came in*). Much similar to the way in which the tagger uses the immediate context of a token for the assignment of the tags, the linguist should include the (whole) context of the tokens when determining the appropriateness of an assigned tag. This is particularly important in relation to the transitivity of lexical verbs.

<i>I <u>see</u> it.</i>	monotransitive: because of the direct object ( <i>it</i> )
<i>I <u>see</u>.</i>	intransitive: because no (direct) object is present

In this manual we discuss cases where the choice of the tag may be problematic under the heading 'notes'.

## 2. MAJOR WORDCLASSES, THEIR SUBCLASSES AND FEATURES

In this part of the manual each of the major wordclasses is defined or described and an inventory is provided of the possible combinations of the wordclasses and features.<sup>1</sup> The major wordclasses are discussed in alphabetical order.

### 2.1 Adjective

Adjectives receive the wordclass label ADJ. Additional feature information relates to subclass (general or ordinal) and form (positive, comparative or superlative; -edp or -ingp; nominal plural). The following tags apply:

ADJ(ge,pos)  
 ADJ(ge,pos,edp)  
 ADJ(ge,pos,ingp)  
 ADJ(ge,comp)  
 ADJ(ge,sup)

ADJ(ord)  
 ADJ(ord,nomplu)

Note that it is only with general adjectives that the form features 'pos', 'comp', and 'sup' are required. The features 'edp' and 'ingp' only occur with positive forms of (general) adjectives. The feature 'nomplu' occurs only in combination with an ordinal adjective.

#### 2.1.1 General adjective

The subclass of general adjectives is an open class. Apart from 'regular' adjectives such as *old*, *impressive*, *tiresome*, *awkward*, etc., this subclass includes 'edp' and 'ingp' adjectives. For example,

<i><u>packed</u> trains</i>	ADJ(ge,pos,edp)
<i><u>stacked</u> boxes</i>	ADJ(ge,pos,edp)
<i>a <u>threatening</u> look</i>	ADJ(ge,pos,ingp)
<i>a <u>dying</u> man</i>	ADJ(ge,pos,ingp)

The tags ADJ(ge,pos,ingp) and ADJ(ge,pos,edp) only apply to adjectives that are deverbal, as in the examples above. Denominal adjectives such as *skilled*, *talented*, *good-hearted* are not 'edp' adjectives.

<i><u>skilled</u> trainees</i>	ADJ(ge,pos)
<i><u>talented</u> singers</i>	ADJ(ge,pos)
<i><u>good-hearted</u> colleagues</i>	ADJ(ge,pos)

While non-gradable adjectives receive the default feature 'pos', gradable adjectives are tagged ADJ(ge,pos), ADJ(ge,comp) or ADJ(ge,sup), depending on whether they have a positive, comparative or superlative form. For example,

<i>The <u>old</u> houses in the village</i>	ADJ(ge,pos)
<i>A more <u>difficult</u> task</i>	ADJ(ge,comp)
<i>The <u>easier</u> thing to do</i>	ADJ(ge,comp)
<i>The <u>saddest</u> story of all</i>	ADJ(ge,sup)

<sup>1</sup> For abbreviations of wordclasses and features, see Appendix 1.

## Notes

- An adjective followed by a common noun, where the adjective does not modify the concept referred to by the noun, but instead creates a subclass within the possible referents of the concept expressed by the noun, is tagged as in the same way as other premodifying adjectives. Compare:

*'nervous system* ADJ(ge,pos) N(sing)  
*nervous 'teacher* ADJ(ge,pos) N(sing)

Other examples are: *common sense, surgical spirit, tangible assets, industrial worker, polar bear.*

- When the adjective is preceded by an intensifying adverb (e.g. *more* or *most*), it has its positive form and should thus be tagged as ADJ(ge,pos):

*He was more persuasive.* ADJ(ge,pos)

- Names of materials, such as *steel, stone, iron, cotton, stone* are tagged as adjectives:

*steel knives* ADJ(ge,pos)  
*stone walls* ADJ(ge,pos)  
*the iron curtain* ADJ(ge,pos)

- General adjectives which function as head of the noun phrase are tagged as nominal adjectives.

*These young people are quite energetic.* ADJ(ge,pos)  
*The young have little idea of what's going on in the world.* NADJ(pos)

- To distinguish 'ingp' and 'edp' adjectives from lexical verb participles: it should be possible to premodify the 'ing' or 'edp' adjective by means of an intensifier (e.g. *very, highly*).

*an isolated village* ADJ(ge,pos,edp)  
*a very surprising present* ADJ(ge,pos,ingp)

- 'edp' forms in premodifying position are always ADJ(ge,pos,edp), even if they do not admit an intensifier:

*an escaped prisoner* ADJ(ge,pos,edp)  
*a fallen angel* ADJ(ge,pos,edp)  
*a born leader* ADJ(ge,pos,edp)

- 'ingp' forms in premodifying position

– are always ADJ(ge,pos,ingp), (even if not modifiable by an intensifier) if the modified noun stands in a subject relation to the *-ing* form:

*a dancing 'girl* ADJ(ge,pos,ingp)  
*barking 'dogs* ADJ(ge,pos,ingp)  
*passing 'cars* ADJ(ge,pos,ingp)  
*rising 'sun* ADJ(ge,pos,ingp)

– if the relation between *-ing* form and modified noun is different, it is not an ADJ(ge,pos,ingp). Instead the *-ing* form is treated as a noun:

a 'dancing master	N(sing)
'racing cars	N(sing)
a 'calculating machine	N(sing)
a 'stumbling block	N(sing)

Note that such combinations have uneven stress. Compare:

'dancing master	N(sing) N(sing)
'dancing 'girl (girl who is dancing)	ADJ(ge,pos) N(sing)

- -ing and -ed forms of adjectives that are introduced by the negative prefix *un-* or *in-* are not tagged with 'ingp' or 'edp':

<u>uncompromising</u> measures	ADJ(ge,pos)
<u>untransformed</u> appearances	ADJ(ge,pos)
<u>inexperienced</u> workers	ADJ(ge,pos)

### Multi-token adjective

When it is hardly possible or senseless to tag one or both items of compound adjectives separately, they are tagged as multi-token units. Also, when hyphenated forms of the compound adjectives exist they should be tagged as multi-token units. Examples are:

1. hyphenation exists:

a <u>dark looking</u> man ( <i>dark-looking</i> )	ADJ(ge,pos):1/2 ADJ(ge,pos):2/2
<u>well off</u> citizens ( <i>well-off</i> )	ADJ(ge,pos):1/2 ADJ(ge,pos):2/2
<u>hands on</u> exercises ( <i>hands-on</i> )	ADJ(ge,pos):1/2 ADJ(ge,pos):2/2
<u>would be</u> champions ( <i>would-be</i> )	ADJ(ge,pos):1/2 ADJ(ge,pos):2/2
<u>pent up</u> feelings ( <i>pent-up</i> )	ADJ(ge,pos):1/2 ADJ(ge,pos):2/2
<u>well meaning</u> compliments ( <i>well-meaning</i> )	ADJ(ge,pos):1/2 ADJ(ge,pos):2/2
<u>to and fro</u> movements	ADJ(ge,pos):1/3 ADJ(ge,pos):2/3 ADJ(ge,pos):3/3

2. separate tagging is hardly possible or senseless:

<u>goose pimply</u> skins	ADJ(ge,pos):1/2 ADJ(ge,pos):2/2
<u>Porto Rican</u> inhabitants	ADJ(ge,pos):1/2 ADJ(ge,pos):2/2
a <u>cash and carry</u> society	ADJ(ge,pos):1/3 ADJ(ge,pos):2/3 ADJ(ge,pos):3/3

3. loan expressions:

<i>à la carte</i>	<i>de jure</i>	<i>par excellence</i>
<i>a posteriori</i>	<i>de luxe</i>	<i>per capita</i>
<i>a priori</i>	<i>de rigueur</i>	<i>per diem</i>
<i>ad hoc</i>	<i>de rigueur</i>	<i>prima facie</i>
<i>all in</i>	<i>ex cathedra</i>	<i>pro forma</i>
<i>all right</i>	<i>ex gratia</i>	<i>pro rata</i>
<i>bona fide</i>	<i>ex officio</i>	<i>sub judice</i>
<i>comme il faut</i>	<i>ex parte</i>	<i>teensy weensy</i>
<i>compos mentis</i>	<i>ex post facto</i>	<i>teeny weeny</i>
<i>cordon bleu</i>	<i>fin de siècle</i>	<i>ultra vires</i>
<i>da capo</i>	<i>hors de combat</i>	<i>viva voce</i>
<i>de facto</i>	<i>non compos mentis</i>	

**Note**

- Of the following premodifying strings the tokens should be tagged separately:

*a sealed off room, small glassed steel containers, low flying aircraft, factor analytic studies, achievement oriented behaviour.*

**2.1.2 Ordinal adjective**

The subclass of ordinal adjectives is a closed class. Its members are:

<i>additional</i>	<i>following</i>	<i>next</i>	<i>previous</i>
<i>another</i>	<i>former</i>	<i>other</i>	<i>prior</i>
<i>certain</i>	<i>further</i>	<i>others</i>	<i>same</i>
<i>extra</i>	<i>latter</i>	<i>penultimate</i>	<i>subsequent</i>
<i>final</i>	<i>last</i>	<i>preceding</i>	

There are no multi-token ordinal adjectives.

NB1: *others* is only used nominally.

NB2: *another* is an exceptional case because it is a contraction of an article with an ordinal adjective.

**Class characteristics**

Syntactic characteristics of the subclass of ordinal adjectives are the following:

1. ordinal adjectives are adjective-like in that they can occur in premodifying position in the noun phrase. Note that some can also occur as head of the noun phrase
2. they can both precede and follow cardinal numerals;
3. they cannot be used together with other ordinal adjectives, or with ordinal numerals.
4. they can precede, but not follow general adjectives.

From a semantic point of view, ordinal adjectives can be characterised as having an ordering function.

**Examples**

<i><u>Certain</u> choices</i>	ADJ(ord)
<i><u>Other</u> options</i>	ADJ(ord)
<i><u>Another</u> man</i>	ADJ(ord)
<i>An <u>additional</u> three people came in.</i>	ADJ(ord)
<i>A <u>further</u> two instances</i>	ADJ(ord)
<i>The <u>same</u> four people.</i>	ADJ(ord)

**Notes**

- Ordinal adjectives such as *penultimate*, *certain*, *another*, *other*, *same* can function as heads of noun phrases, like general adjectives. The latter receive the tag NADJ (nominal adjective) in that function, whereas ordinal adjectives do not. Instead, they are tagged as ADJ(ord).

<i>The <u>young</u> are dancing.</i>	NADJ(pos)
<i>The <u>same</u> applies for you.</i>	ADJ(ord)

See also ordinal numerals (section 2.8.2).

- The feature ADJ(ord,nomplu) is used for *others*, which is used nominally only, with a plural form:

*Others will have to be sold to other institutions.*      ADJ(ord,nomplu)

## 2.2 Adverb

Adverbs receive the wordclass label ADV. Additional feature information relates to subclass (general, connective, negative, phrasal or *wh*), and form (positive, comparative or superlative). The following tags apply:

ADV(ge,pos)  
ADV(ge,comp)  
ADV(ge,sup)

ADV(connec)  
ADV(neg)  
ADV(phras)  
ADV(wh)

Note that it is only with general adverbs that a second feature relating to their form is required.

### 2.2.1 General adverb

The general adverbs form a default category; its members are all adverbs that do not belong to the other adverb subclasses. This implies that the following traditional subclasses also belong to the subclass of general adverbs: additive, assertive, emphatic, exclusive, intensifying, nonassertive, particularising and qualifying adverbs.

The subclass of general adverbs is an open class. With non-gradable general adverbs the feature 'pos' is associated. For gradable general adverbs one of the following features can be selected: 'pos' (positive), 'comp' (comparative), 'sup' (superlative).

#### Examples

<i>This car runs very <u>fast</u>.</i>	ADV(ge,pos)
<i>You should get up <u>earlier</u>.</i>	ADV(ge,comp)
<i>Which one do you like <u>most</u>?</i>	ADV(ge,sup)
<i><u>Yesterday</u> he arrived in Lisbon.</i>	ADV(ge,pos)
<i>Why did he <u>ever</u> refuse the offer?</i>	ADV(ge,pos)
<i>He <u>alone</u> can help us.</i>	ADV(ge,pos)
<i>He is still <u>but</u> a child.</i>	ADV(ge,pos)
<i>I was <u>merely</u> trying to help.</i>	ADV(ge,pos)
<i>She did <u>fairly</u> well.</i>	ADV(ge,pos)
<i>They were <u>nearly</u> there.</i>	ADV(ge,pos)
<i>She was <u>the</u> more surprised when he told her he was leaving.</i>	ADV(ge,pos)
<i>It was <u>well-nigh</u> impossible.</i>	ADV(ge,pos)
<i>It couldn't get <u>any</u> worse.</i>	ADV(ge,pos)
<i>She was <u>mainly</u> interested in the earlier species.</i>	ADV(ge,pos)
<i>The lights went out <u>bang</u> in the middle of the performance.</i>	ADV(ge,pos)
<i>He ran <u>bump</u> into a tree.</i>	ADV(ge,pos)
<i>The bullet went <u>right</u> through his heart.</i>	ADV(ge,pos)

#### Notes

- General adverbs can function as noun phrase premodifiers, as in:

*The then president resigned.*      ADV(ge,pos)

Or as noun phrase postmodifiers, as in:

*The examples above are from a corpus.* ADV(ge,pos)

- *there* can also be existential *there*. It may co-occur with the general adverb in the same sentence:

*There are many people who desperately want to go.* EXTHERE  
*There weren't many people there.* EXTHERE  
*There weren't many people there.* ADV(ge,pos)

- *any* can also be a pronoun. Compare:

*Can't you wait any longer?* ADV(ge,pos)  
*I don't want any tea.* PRON(quant)

- *but* can be an adverb, a preposition and a coordinator. Compare:

*He ate all chocolates but two.* PREP(ge)  
*He doesn't mind, but doesn't find it ideal either.* CONJUNC(coord)  
*He was but five years old.* ADV(ge,pos)

- *as* can be an adverb, a subordinator and a preposition. Compare:

*He was as tall as Peter.* ADV(ge,pos)  
*He was careful, as he knew that one mistake would be fatal.* CONJUNC(subord)  
*As a teacher, he earned a comfortable salary.* PREP(ge)

### Multi-token general adverb

While the subclass of general adverbs is an open class, the set of multi-token items in this subclass is a limited one. It includes the following items:

<i>a fortiori</i>	<i>at once</i>	<i>fore and aft</i>
<i>à la carte</i>	<i>at present</i>	<i>hard by</i>
<i>à la mode</i>	<i>au fond</i>	<i>hors de combat</i>
<i>a posteriori</i>	<i>bona fide</i>	<i>in extremis</i>
<i>a priori</i>	<i>by and by</i>	<i>in general</i>
<i>ab initio</i>	<i>by and large</i>	<i>in loco parentis</i>
<i>ad hoc</i>	<i>by far</i>	<i>in particular</i>
<i>ad infinitum</i>	<i>da capo</i>	<i>in situ</i>
<i>ad interim</i>	<i>de facto</i>	<i>in toto</i>
<i>ad lib(itum)</i>	<i>de jure</i>	<i>inter alia</i>
<i>ad nauseam</i>	<i>early on</i>	<i>ipso facto</i>
<i>ad valorem</i>	<i>en claire</i>	<i>kind of</i>
<i>all of a sudden</i>	<i>en famille</i>	<i>later on</i>
<i>all the more</i>	<i>en masse</i>	<i>mutatis mutandis</i>
<i>ante meridiem</i>	<i>en passant</i>	<i>nem con</i>
<i>a.m.</i>	<i>en plein air</i>	<i>nip and tuck</i>
<i>am</i>	<i>en route</i>	<i>of course</i>
<i>as such</i>	<i>ex cathedra</i>	<i>on average</i>
<i>at all</i>	<i>ex gratia</i>	<i>on the whole</i>
<i>at first</i>	<i>ex officio</i>	<i>per annum</i>
<i>at last</i>	<i>ex parte</i>	<i>per capita</i>
<i>at least</i>	<i>ex post facto</i>	<i>per cent</i>

*per diem*  
*per se*  
*post meridiem*  
*p.m.*  
*pm*  
*prima facie*

*pro forma*  
*pro rata*  
*pro tem(pore)*  
*right away*  
*sine die*  
*sort of*

*sotto voce*  
*sub rosa*  
*to and fro*  
*tout ensemble*  
*viva voce*

### Note

- *sort of* can be a multi-token general adverb. For example,

This sort of brings us back to the beginning.

ADV(ge,pos):1/2 ADV(ge,pos):2/2

### 2.2.2 Connective adverb

The subclass of connective adverbs is a closed class (cf. CGEL 8.116: additive adverbs). Its members are:

*also*  
*and*  
*anyhow*  
*anyway*  
*besides*  
*both*  
*but*  
*e.g.*  
*eg*  
*either*  
*finally*  
*first*  
*firstly*  
*further*  
*furthermore*  
*hence*

*however*  
*i.e.*  
*ie*  
*including*  
*indeed*  
*last*  
*lastly*  
*like*  
*more*  
*moreover*  
*namely*  
*neither*  
*nevertheless*  
*next*  
*nonetheless*  
*nor*

*notwithstanding*  
*or*  
*plus*  
*second*  
*secondly*  
*say*  
*so*  
*still*  
*therefore*  
*third*  
*thirdly*  
*though*  
*too*  
*viz.*  
*viz*  
*yet*

Multi-token units:

*above all*  
*after all*  
*all in all*  
*all the same*  
*as follows*  
*as well*  
*at any rate*  
*by contrast*  
*by the by*  
*by the bye*  
*by the way*  
*even so*  
*first of all*  
*for all that*  
*for another*

*for another thing*  
*for example*  
*for instance*  
*for one*  
*for one thing*  
*in actual fact*  
*in addition*  
*in all*  
*in fact*  
*in particular*  
*in short*  
*in sum*  
*in the first place*  
*in the nth place*  
*in the main*

*last of all*  
*not to mention*  
*on the contrary*  
*on the one hand*  
*on the other*  
*on the other hand*  
*on the whole*  
*once again*  
*once more*  
*so to speak*  
*so to say*  
*that is*  
*that is to say*  
*what's more*

Note that while the class of connective adverbs is a closed class, the full set of items cannot be listed here since the set of items that answer the frame 'in the *n*th place' (where *n* can be any number) is not enumerable.



### Class characteristics

A connective adverb establishes a logical relation between the current utterance and (a/the) previous one(s), a function they have in common with coordinators. Unlike coordinators, however, their position in the utterance is not fixed (e.g. *as well, too, so to speak*).

### Examples

*She's a sort of angel, so to speak.*

ADV(connect):1/3 ADV(connect):2/3 ADV(connect):3/3

*What's more: he's barking mad.*

ADV(connect):1/3 ADV(connect):2/3 ADV(connect):3/3

*On the one hand she's handsome, on the other hand she's a bit of a cat*

ADV(connect):1/3 ADV(connect):2/3 ADV(connect):3/3

### Notes

- The following are tagged as general adverbs: *actually, again, alternatively, altogether, conversely, else, equally, incidentally, instead, likewise, now, only, otherwise, overall, rather, then, thus*.
- Of the following expressions the tokens are tagged separately: *as a consequence, as a result, as a rule, at all events, at any event, at the same time, by comparison, by contrast, by the same token, by way of comparison, by way of contrast, for a start, in any case, in any event, in comparison, in conclusion, in consequence, in contrast, in other words, in that case, in the same place, in the same way*.
- If coordinators are sentence-initial, they should be tagged as connective adverbs

*And who have we got here?*

ADV(connect)

*Or do you know the solution?*

ADV(connect)

- *neither, either* and *both* can also be pronouns. Compare:

*Neither book was expensive.*

PRON(neg)

### 2.2.3 Negative adverb

The subclass of negative adverbs is a closed class. Its members are:

*barely*

*no*

*scarcely*

*hardly*

*not*

*seldom*

*never*

*nowhere*

*nevermore*

*rarely*

There are no multi-token negative adverbs.

### Class characteristics

The adverbs have a negative meaning component and, when sentence-initial, cause partial inversion.

### Examples

*They had never seen anything like this.*

ADV(neg)

*No sooner had he said it than he wished he hadn't.*

ADV(neg)

*They did not hand in their papers.*

ADV(neg)

**Notes**

- *No* can also be a negative pronoun, or form part of a multi-token negative pronoun. Compare:

*No* sooner had he said it than he wished he hadn't. ADV(neg)  
*No* person can do such a thing. PRON(neg)  
*No one* knows what the future will look like. PRON(neg):1/2 PRON(neg):2/2

**2.2.4 Phrasal adverb**

The subclass of phrasal adverbs is a closed class (cf. CGEL 16.3f; 16.6). Among its members are:

<i>about</i>	<i>away</i>	<i>in</i>	<i>out</i>	<i>through</i>
<i>along</i>	<i>back</i>	<i>off</i>	<i>over</i>	<i>up</i>
<i>around</i>	<i>down</i>	<i>on</i>	<i>round</i>	

There are no multi-token phrasal adverbs.

**Class characteristics**

A phrasal adverb forms part of a phrasal verb. Instead of tagging a phrasal verb as a multi-token unit, each of the constituent parts is tagged separately.

**Examples**

*He gave up.* ADV(phras)  
*He looked it up.* ADV(phras)

**Notes**

- The major difference between phrasal prepositions and phrasal adverbs is brought out by pronominalization of the object NP. Consider the following examples:

*The student looked up the word in the dictionary.* VB(lex,montr,past) ADV(phras)  
*The student looked at the word in the dictionary.* VB(lex,montr,past) PREP(phras)

Pronominalization of the object NP results in:

*The student looked it up in the dictionary.* ADV(phras)  
 \**The student looked up it in the dictionary.*  
*The student looked at it in the dictionary.* PREP(phras)  
 \**The student looked it at in the dictionary.*

**2.2.5 Wh-adverb**

The subclass of *wh*-adverbs is a closed class. Its members are:

<i>how</i>	<i>whereat</i>	<i>whereunto</i>
<i>however</i>	<i>whereby</i>	<i>whereupon</i>
<i>howsoever</i>	<i>wherefore</i>	<i>wherever</i>
<i>when</i>	<i>wherein</i>	<i>whither</i>
<i>whence</i>	<i>whereof</i>	<i>why</i>
<i>whenever</i>	<i>whereon</i>	
<i>whensoever</i>	<i>wheresoever</i>	
<i>where</i>	<i>whereto</i>	

There are no multi-token *wh*-adverbs.

### Class characteristics

Adverbs belonging to this class are all adverbs beginning with *wh-* as well as *how*, *however*, *howsoever*. This implies that interrogative, relative as well as intensifying *wh*-adverbs are included.

### Examples

<i>How</i> did he do it?	ADV(wh)
<i>When</i> does the train leave?	ADV(wh)
<i>Where</i> are you going?	ADV(wh)
She never knew <i>why</i> he left.	ADV(wh)

## 2.3 Article

Articles receive the wordclass label ART. Additional information relates to their reference (definite or indefinite). The following tags apply:

ART(def)  
ART(indef)

The class of articles is a closed class. Its members are: *a*, *an*, *the* and the archaic form *ye*.

### Examples

<i>The</i> man came in.	ART(def)
<i>The</i> one I know is much taller.	ART(def)
<i>The</i> rich in this country have become richer over <i>the</i> past five years.	ART(def)
I know he's always been <i>a</i> hero.	ART(indef)
Would you like <i>a</i> cup of tea?	ART(indef)

### Notes

- the* can also be an adverb, e.g.

<i>The</i> more you know about this, <i>the</i> less you like it.	ADV(ge,pos)
<i>The</i> longer they waited <i>the</i> colder they got.	ADV(ge,pos)
<i>The</i> sooner you call me, <i>the</i> better it is.	ADV(ge,pos)

## 2.4 Conjunction

Conjunctions receive the wordclass label CONJUNC. Additional feature information relates to subclass (co-ordinating or subordinating). The following tags apply:

CONJUNC(coord)  
CONJUNC(subord)

### Class characteristics

Conjunctions are typically 'link' words in the constituents in which they occur; this implies that they can never realize another function in the standard structure of their constituent.

### 2.4.1 Co-ordinating conjunction

The subclass of co-ordinating conjunctions is a closed class. Its members are:

&	<i>and/or</i>	<i>nor</i>	<i>yet</i>
/	<i>but</i>	<i>plus</i>	
<i>and</i>	<i>for</i>	<i>or</i>	



<i>cos</i>	<i>since</i>	<i>whenever</i>
<i>directly</i>	<i>so</i>	<i>where</i>
<i>ere</i>	<i>than</i>	<i>wherever</i>
<i>forasmuch</i>	<i>that</i>	<i>whereas</i>
<i>if</i>	<i>tho'</i>	<i>whereupon</i>
<i>immediately</i>	<i>though</i>	<i>whether</i>
<i>lest</i>	<i>till</i>	<i>while</i>
<i>like</i>	<i>unless</i>	<i>whilst</i>
<i>now</i>	<i>until</i>	
<i>once</i>	<i>when</i>	

Multi-token units:

<i>as far as</i>	<i>for fear</i>	<i>insomuch as</i>
<i>as if</i>	<i>for fear that</i>	<i>in so much as</i>
<i>as long as</i>	<i>in as far as</i>	<i>in that</i>
<i>as soon as</i>	<i>insofar as</i>	<i>now that</i>
<i>as though</i>	<i>inasmuch as</i>	<i>once that</i>
<i>for all as</i>	<i>in as much as</i>	<i>rather than</i>
<i>for all that</i>	<i>in case</i>	<i>save that</i>
<i>for as much</i>	<i>in order that</i>	<i>so far as</i>
<i>forasmuch as</i>	<i>insofar as</i>	<i>so long as</i>
<i>for as much as</i>	<i>in so far as</i>	<i>so that</i>
<i>forasmuch that</i>	<i>insofar that</i>	<i>sooner than</i>
<i>for as much that</i>	<i>in so far that</i>	<i>whether or not</i>

### Class characteristics

Syntactically, subordinating conjunctions can be characterised as follows:

1. a subordinating conjunction introduces a subclause, i.e. a clause functioning as constituent of a superordinate clause or of a phrase.
2. the subordinating conjunction is clause-initial.

### Examples

<i>The boy left, <u>because</u> it was getting dark.</i>	CONJUNC(subord)
<i>I'll call you, <u>if</u> you give me your telephone number.</i>	CONJUNC(subord)
<i>He smiled, <u>as if</u> he knew all her secrets.</i>	
	CONJUNC(subord):1/2 CONJUNC(subord):2/2

### Notes

- If *whether or not* occurs as an interrupted string (i.e. when the constituent parts are not adjacent), *whether* is tagged as CONJUNC(subord), while *or* is tagged as CONJUNC(coord) and *not* as PROFM(so,clause).

*Whether he agrees or not, we are going through with it.*  
 CONJUNC(subord) CONJUNC(coord)PROFM(so,clause)

- The combinations *even though* and *even if* are not multi-token conjunctions but sequences consisting of adverbs followed by subordinating conjunctions.

<i>even though</i>	ADV(ge,pos) CONJUNC(subord)
<i>even if</i>	ADV(ge,pos) CONJUNC(subord)

## 2.5 *Existential there*

Existential *there* receives the wordclass label EXTHERE. The wordclass is closed and consists only of the word *there*. There is no additional feature information. The tag that applies is

EXTHERE

### Class characteristics

The wordclass label is assigned to unstressed *there* in existential sentences, which are typically of the form: *there + be + indefinite noun phrase*. In some cases a verb different from *be* may occur.

### Examples

<i><u>There</u> are a few things I would like to know.</i>	EXTHERE
<i>Is <u>there</u> anything I can do?</i>	EXTHERE

### Notes

- Usually in clauses and sentences with existential *there* the verb (phrase) is intransitive:

*There's a man at the door.*  
*There could be a lot of money in this.*  
*There came a moment when he felt desperate.*  
*There still remains one problem.*  
*There comes a point when the reason in mathematics begins to fail.*  
*There's also included in this class a lot of other things.*

- There* can also be a general adverb. Existential *there* and the general adverb *there* may co-occur in the same sentence.

<i>It isn't neither here nor <u>there</u>.</i>	ADV(ge,pos)
<i><u>There</u> is nothing here.</i>	EXTHERE
<i><u>There's</u> nothing of interest for us <u>there</u>.</i>	EXTHERE
	ADV(ge,pos)

## 2.6 *Noun*

Nouns receive the wordclass label N. Additional feature information relates to their form (singular, plural, number), while it is only with collective nouns that a second feature is assigned indicating their subclass membership. The following tags apply:

N(sing)  
 N(plu)  
 N(number)  
  
 N(sing,collect)  
 N(plu,collect)

Note that no distinction is made between 'common' and 'proper' nouns; nor do we distinguish between 'count' and 'non-count' nouns.

### Class characteristics

Most nouns can take two inflectional suffixes, one to mark number (the plural) and one to mark case (the genitive). In addition, a great many nouns can be identified on the basis of typical derivational suffixes such as *-age*, *-ance*, *-ation*, *-ess*, *-hood*, *-ism*, *-ness*, etc.

### The form features 'singular', 'plural', and 'number'

As a general rule, the form of the noun determines its number: if the noun has the singular form, it is tagged as 'singular'; if it has the plural form it is tagged as 'plural'; when singular and plural have the same form, the feature 'number' is used. For example,

<i>house</i> N(sing)	<i>houses</i> N(plu)	<i>sheep</i> N(number)
<i>Monday</i> N(sing)	<i>Mondays</i> N(plu)	<i>deer</i> N(number)
<i>stimulus</i> N(sing)	<i>analyses</i> N(plu)	

There are two exceptions to the general rule above:

- I. A few groups of words form an exception to the rule that a plural form always entails the 'plural' tag:

1. names of academic or other disciplines ending in suffix *-ics* are tagged as 'singular'; for example,

<i>mathematics</i>	<i>politics</i>	<i>economics</i>
<i>statistics</i>	<i>ethics</i>	<i>gymnastics</i>
<i>physics</i>	<i>linguistics</i>	<i>phonetics</i>
<i>athletics</i>	<i>classics</i>	

Compare:

<i>statistics</i> ('branch of science')	N(sing)
<i>statistic</i> ('single number, measurement')	N(sing)
<i>statistics</i> ('collected numbers, measurements')	N(plu)
<i>acoustics</i> ('the scientific study of sound')	N(sing)
<i>acoustics</i> ('the qualities of a place in relation to sound')	N(plu)
<i>acoustics</i> ('the qualities of sounds picked up by one's ear')	N(plu)

2. names of games are tagged as 'singular', irrespective of their plural form; for example,

<i>checkers</i>	<i>dominoes</i>	<i>fives</i>
<i>draughts</i>	<i>drafts</i>	<i>ninepins</i>

In case these wordforms have other senses which allow the singular form, their singular form should be tagged as singular, their plural form as plural.

Compare:

<i>darts</i> ('game')	N(sing)
<i>dart</i> ('small pointed object')	N(sing)
<i>darts</i> ('small pointed objects')	N(plu)

3. names of diseases are tagged as singular; for example,

<i>measles</i>	<i>rickets</i>
<i>mumps</i>	<i>shingles</i>

4. other

<i>Brussels</i>	N(sing)
<i>news</i>	N(sing)

II. The following are examples of nouns that are always plural, irrespective of their form:

<i>accoutrements</i>	<i>bourgeois**</i>	<i>kinfolk</i>
<i>accouterments</i>	<i>breeches</i>	<i>kinsfolk</i>
<i>arms*</i>	<i>bristols</i>	<i>menfolk</i>
<i>battlements</i>	<i>britches</i>	<i>military</i>
<i>bed-clothes</i>	<i>clergy</i>	<i>personnel</i>
<i>bedclothes</i>	<i>folk</i>	<i>police</i>
<i>belongings</i>	<i>gentlefolk</i>	<i>populace</i>
<i>bifocals</i>	<i>kin</i>	<i>townsfolk</i>
<i>binoculars</i>	<i>kindred</i>	<i>tradesfolk</i>

\* heraldry  
\*\* social class

as well as some proper names, such as

<i>the Alps</i>	N(plu)
<i>the Falklands</i>	N(plu)

### 'compound' nouns

In sequences of words that might be looked upon as constituting 'compound' nouns, the words of the sequence are tagged separately whenever it is feasible to do so, in accordance with the general policy with respect to candidate multi-token units. For example,

<i>Australian Rules football</i>	ADJ(ge,pos) N(plu) N(sing)
<i>electric shock therapy</i>	ADJ(ge,pos) N(sing) N(sing)
<i>high water mark</i>	ADJ(ge,pos) N(sing) N(sing)
<i>forty winks</i>	NUM(card,sing) N(sing)
<i>first offender</i>	NUM(ord,sing) N(sing)
<i>notary public</i>	N(sing) ADJ(ge,pos)
<i>heir apparent</i>	N(sing) ADJ(ge,pos)
<i>bedroom door</i>	N(sing) N(sing)
<i>water melon</i>	N(sing) N(sing)

Separate tagging of the tokens in a lexical string also takes place when the semantic relation between the adjective and noun is more complex:

<i>dead march</i>	ADJ(ge,pos) N(sing)
<i>poor box</i>	ADJ(ge,pos) N(sing)
<i>poor law</i>	ADJ(ge,pos) N(sing)
<i>funny farm</i>	ADJ(ge,pos) N(sing)
<i>sitting duck</i>	ADJ(ge,pos,ingp) N(sing)
<i>red tape</i>	ADJ(ge,pos) N(sing)
<i>fixed odds</i>	ADJ(ge,pos,edp) N(plu)



*American football*  
*biological warfare.*

ADJ(ge,pos) N(sing)  
ADJ(ge,pos) N(sing)

Only if the tag for one of the words in the sequence is uncertain or would be inconsequential, are such sequences treated as 'compounds'. This is, for example, the case when:

1. the compound noun does not have a nominal head (e.g. *cash and carry*)
2. the compound noun contains a token which is only used in the compound (e.g. *cos lettuce*), or it contains a token for which it is impossible to determine its wordclass (e.g. *cover note, alma mater*).

#### **Examples of compound nouns which do not have a nominal head**

The following compounds are all ditto-tagged N(sing).

*about turn*

*deep freeze*

*rock and roll*

*all clear*

*fly half*

*rub down*

*bubble and squeak*

*hue and cry*

*stand-off half*

*cash and carry*

*leg bye*

*cat-o'-nine tails*

*plus fours*

#### **Examples of absent sense compounds**

The following are tagged as singular compound nouns because the wordclass of at least one of the tokens is unclear (since the token is used only in the compound).

*auld lang syne*

*finnan haddie*

*marram grass*

*Bren gun*

*finnan haddock*

*milch cow*

*conger eel*

*guelder rose*

*ouija board*

*cos lettuce*

*ichneumon fly*

*pease pudding*

*dak bungalow*

*joss stick*

*submachine gun*

*deckle edge*

*lateen sail*

*sulfa drug*

*decree nisi*

*lese majesty*

*sulpha drug*

#### **Examples of other compounds containing elements whose wordclass is uncertain**

The following compounds are all ditto-tagged N(sing) because it is uncertain whether the first token should be tagged as a verb or as a noun.

*crash helmet*

*bubble gum*

*hobble skirt*

*fancy man*

*clasp knife*

*hush money*

*call sign*

*brake wheel*

*goggle box*

*catch crop*

*boom town*

*jam session*

*drop hammer*

*close season*

*flick knife*

*drift ice*

*crash landing*

*dispatch rider*

*cover note*

#### **Examples of compounds of foreign origin**

Again, the following are tagged as compounds. All are ditto-tagged N(sing), except for *finest herbes* which is ditto-tagged N(plu).

*agent provocateur*

*chargé d'affaires*

*coup de grace*

*alma mater*

*chef d'oeuvre*

*crêpe suzette*

*alter ego*

*chow chow*

*curriculum vitae*

*au pair*

*cordon bleu*

*entente cordiale*

*bar mitzvah*

*corpus delicti*

*finest herbes*

*bass clef*

*coup d'état*

*foie gras*

<i>grand mal</i>	<i>per cent</i>	<i>prima donna</i>
<i>magnum opus</i>	<i>petit four</i>	<i>quid pro quo</i>
<i>Mardi gras</i>	<i>petit mal</i>	<i>raison d'être</i>
<i>nom de plume</i>	<i>pièce de résistance</i>	<i>sans serif</i>
<i>non sequitur</i>	<i>plat du jour</i>	<i>shish kebab</i>
<i>obiter dictum</i>	<i>porte cochere</i>	<i>sine qua non</i>
<i>objet d'art</i>	<i>prima ballerina</i>	

### Abbreviations and acronyms

Abbreviated forms should be tagged as their full forms. Examples are:

<i>pop art</i>	<i>(popular art)</i>	ADJ(ge,pos) N(sing)
<i>long vac</i>	<i>(long vacation)</i>	ADJ(ge,pos) N(sing)
<i>op art</i>	<i>(optical art)</i>	ADJ(ge,pos) N(sing)
<i>small ad</i>	<i>(small advertisement)</i>	ADJ(ge,pos) N(sing)
<i>classified ad</i>	<i>(classified advertisement)</i>	ADJ(ge,pos,edp) N(sing)

Acronyms:

- a) compound acronyms should be tagged as single tokens. The forms should be tagged as singular nouns, unless they end in plural (lowercase) s

<i>the <u>UN</u></i>	N(sing)
<i>the <u>UNESCO</u></i>	N(sing)
<i>a <u>POW</u></i>	N(sing)
<i>two <u>POWs</u></i>	N(plu)
<i>the <u>UK</u></i>	N(sing)

- b) acronyms consisting of separate letters: of acronyms consisting of letters separated by spaces each letter should be tagged as a single noun, no matter whether it stands for a (singular or plural) noun or adjective or other wordclass. Examples are:

<i><u>I</u> <u>D</u> card</i>	N(sing) N(sing)
<i><u>P</u> <u>T</u> boat</i>	N(sing) N(sing)
<i><u>U</u> <u>K</u></i>	N(sing) N(sing)
<i><u>U.</u> <u>K.</u></i>	N(sing) N(sing)

### Notes

- Even if the noun has plural meaning, it is tagged as singular when the form is singular. Compare:

<i>10 <u>metres</u></i>	N(plu)	<i>50 <u>lions</u></i>	N(plu)
<i>10 <u>metre</u></i>	N(sing)	<i>50 <u>lion</u></i>	N(sing)

Some nouns have singular and plural forms, where the plural form has both singular and plural meaning. Their form determines which feature should be assigned. An example is:

<i>assize</i>	N(sing)	<i>assizes</i>	N(plu)
---------------	---------	----------------	--------

Some nouns have singular and plural forms, where the singular form has both singular and plural meaning. Again, the form determines which feature should be assigned. Examples are:

1. Words meaning 'ship', ending in *-craft*:

<i>hovercraft</i>	N(sing)	<i>hovercrafts</i>	N(plu)
<i>spacecraft</i>	N(sing)	<i>spacecrafts</i>	N(plu)

2. Game (i.e. animals that can be hunted); examples are:

<i>lion</i>	N(sing)	<i>lions</i>	N(plu)
<i>cod</i>	N(sing)	<i>cods</i>	N(plu)
<i>elephant</i>	N(sing)	<i>elephants</i>	N(plu)

3. Insects or other small animals which cause disease or damage. Examples are:

<i>termite</i>	N(sing)	<i>termites</i>	N(plu)
<i>moth</i>	N(sing)	<i>moths</i>	N(plu)
<i>flea</i>	N(sing)	<i>fleas</i>	N(plu)

4. Weights, measures, currencies, etc. Examples are:

<i>kilometre</i>	N(sing)	<i>kilometres</i>	N(plu)
<i>mile</i>	N(sing)	<i>miles</i>	N(plu)
<i>acre</i>	N(sing)	<i>acres</i>	N(plu)
<i>kilogram</i>	N(sing)	<i>kilograms</i>	N(plu)
<i>tonne</i>	N(sing)	<i>tonnes</i>	N(plu)
<i>litre</i>	N(sing)	<i>litres</i>	N(plu)
<i>gallon</i>	N(sing)	<i>gallons</i>	N(plu)
<i>guilder</i>	N(sing)	<i>guilders</i>	N(plu)

- Deverbal nouns have one or more of the following nominal characteristics:
  - they can be pluralised (e.g. *killing/killings*)
  - they can be preceded by one or more determiners (e.g. *her few writings*)
  - they can have one or more premodifiers (e.g. *a cold-blooded killing*)
  - they can form a compound with a following noun (e.g. *dancing master*).
  - they can have one or more postmodifiers (e.g. *the writing on the wall*).

If a verb-like word is followed by a direct object, it is tagged as a verb:

*I don't like his smoking cigars.* VB(lex,motr,ingp)

- Genitive markers are treated as separate tokens.

*Peter's car* N(sing) GENM

- Compounds consisting of sequences of non-English tokens which together constitute a proper name are tagged as follows: name parts beginning with a capital letter should be tagged as N(sing), no matter what their wordclasses in the other language are; all parts in lowercase should be tagged as MISC(foreign), no matter what their wordclasses in the other language are. Examples are:

*He lived in Broek op Lange Dijk* N(sing) MISC(foreign) N(sing) N(sing)  
*He's a friend of Peter van der Toorn* N(sing) MISC(foreign) MISC(foreign) N(sing)

- Hyphenated word forms are treated as one token. The point is here that there is no deciding which is the 'headword'. In compounds this is usually assumed to be the second element, but this is counter-intuitive in some cases and would lead to chaotic tagging. Tokens that are hyphenated and whose last element is a noun are treated as single

adjectives: *four-wheel drive*, *sickle-cell anaemia*, *internal-combustion engine*, *foot-and-mouth disease*, *out-of-pocket expenses*. This also holds for compounds containing a genitive marker which is hyphenated: *hound's-tooth check*, *bird's-eye view*.

- The tokens of nominalised clauses should be tagged separately, unless these tokens are hyphenated. Compare:

*what do you call it*  
PRON(inter) VB(aux,do,pres) PRON(pers,sing) VB(lex,cxtr,inf) PRON(pres,sing)

*what-do-you-call-it*  
N(sing)

And:

*what's its name*  
PRON(inter) VB(lex,cop,pres,encl) PRON(poss,sing) N(sing)

*what's-its-name*  
N(sing)

- As said at the beginning of this chapter, proper nouns do not constitute a separate category in the ICLE tagset. Proper names receive the regular noun tag:

*John* N(sing)  
*I know two Peters* N(plu)

Proper names which end in s but have singular meaning are tagged as singular. Examples are:

*Miles Davis* N(sing) N(sing)  
*Oliver Sacks* N(sing) N(sing)

- Potential compound proper names are treated in the same way as other sequences, i.e. their elements are tagged as separate tokens, unless the tagging of one of their words is problematic:

*New York* ADJ(ge,pos) N(sing)  
*Peter de Vries* N(sing) MISC(foreign) N(sing)

- Titular abbreviations are tagged as nouns; e.g. *Dr John*, *Mr James*, *Ms Smith*, *Private Peterson*, *Captain Jones*, *Lady Chatterley*, *King Henry VIII*, *Messrs Peterson and Smith*.

- Letters, including abbreviations, are tagged as nouns:

*A level* N(sing)  
*A to Z* N(sing)

- All the tokens of titles are tagged for their wordclasses, unless they are non-English. For non-English words the tag MISC(foreign) is used. Examples are:

*Who's who* PRON(inter) VB(lex,cop,pres,encl) PRON(inter)  
*Trading places* VB(lex,motr,ingp) N(plu)  
*As you like it* CONJUNC(subord) PRON(pers,number) VB(lex,montr,pres)  
PRON(pers,sing)

- As opposed to *Northern, Eastern, Southern, Western*, the tokens *North, East, South, West* are tagged as nouns:

<i>North London</i>	N(sing) N(sing)
<i>East Africa</i>	N(sing) N(sing)
<i>South Africa</i>	N(sing) N(sing)

### 2.6.1 Collective nouns

The subclass of collective nouns is, in principle, a closed class. Its members are:

<i>agency (e.g. travel agency)</i>	<i>fringe (e.g. lunatic fringe)</i>
<i>audience (e.g. studio audience)</i>	<i>games (e.g. Olympic Games)</i>
<i>band (e.g. steel/string band)</i>	<i>generation (e.g. rising generation)</i>
<i>board</i>	<i>government</i>
<i>bracket (e.g. age bracket)</i>	<i>group</i>
<i>brigade (e.g. fire brigade)</i>	<i>group (e.g. age group)</i>
<i>chamber (e.g. chamber of commerce)</i>	<i>group (e.g. ginger group)</i>
<i>class (e.g. lower/upper/middle class)</i>	<i>guard (e.g. coast guard)</i>
<i>college (e.g. electoral college)</i>	<i>guard (e.g. old guard)</i>
<i>commission (e.g. high commission)</i>	<i>militia</i>
<i>committee (e.g. select committee)</i>	<i>network (e.g. old boy network)</i>
<i>company (e.g. stock company)</i>	<i>party (e.g. fatigue/search/stretchers party)</i>
<i>core (e.g. hard core)</i>	<i>service</i>
<i>council (e.g. county/town council)</i>	<i>set (e.g. jet set)</i>
<i>counsel</i>	<i>sex (e.g. the fair/gentle sex)</i>
<i>court</i>	<i>shift (e.g. night shift)</i>
<i>crew (e.g. ground crew)</i>	<i>squad (e.g. bomb disposal squad)</i>
<i>deck (e.g. lower deck)</i>	<i>staff (e.g. general/ground staff)</i>
<i>department (e.g. fire department)</i>	<i>wing (e.g. left/right wing)</i>
<i>force (e.g. task force)</i>	
<i>form (e.g. sixth form)</i>	

They are tagged N(sing,collect) when they are singular in form and N(plu,collect) when they are plural in form.

#### Class characteristics

The class of collective 'common' nouns consists of nouns which have different forms for singular and plural. Their singular or plural form can show either concord or lack of concord with the verb. Semantically, they usually denote groups of people, who can either be looked upon as individuals (in which case there is lack of concord) or as the group constituting a single entity (in which case there is concord):

*The government has/have decided.*  
*The games is/are in Atlanta.*

#### Notes

- Although the class of collective nouns is in principle closed, it can be extended through the addition of collective 'proper' names. Examples are:

<i>Downing Street</i>	<i>Free Church</i>	<i>Home Guard</i>
<i>Eurovision</i>	<i>Gallup</i>	<i>Home Office</i>
<i>Fleet Street</i>	<i>Harley Street</i>	<i>House of Commons</i>
<i>Foreign Office</i>	<i>Hollywood</i>	<i>House of Lords</i>

*Inland Revenue*  
*Interpol*  
*Jockey Club*  
*King's Bench*  
*Ku Klux Klan*  
*Labour Party*  
*Liberal Party*  
*Lloyd's (Bank)*

*Marine Corps*  
*Met (Office)*  
*Metropolitan Police*  
*New Left*  
*Olympic Games*  
*Oxbridge*  
*Peace Corps*  
*Privy Council*

*Red Crescent*  
*Red Cross*  
*Scotland Yard*  
*Security Council*  
*Sinn Fein*  
*Wall Street*  
*Whitehall*

## 2.7 Nominal adjective

Nominal adjectives (cf. CGEL 7.23-26) receive the wordclass label NADJ. Additional feature information relates to form (positive, comparative or superlative), while with (nominal) 'ing' and 'ed' adjectives a second feature indicates the fact that these are deverbal. The following tags apply:

NADJ(pos)  
 NADJ(pos,edp)  
 NADJ(pos,ingp)  
 NADJ(comp)  
 NADJ(sup)

### Class characteristics

The class of nominal adjectives is an open class. It can be characterised as follows:

Syntactically:

1. a nominal adjective is a general adjective which functions as the head of a noun phrase;
2. many nominal adjectives may be modified by intensifying adverbs (*very*, *more*, etc.);
3. a nominal adjective always requires a determiner.

Morphologically:

1. a nominal adjective does not take the plural form, although it often is plural semantically. Neither does it allow genitive inflections;
2. nominal adjectives may have the positive (or absolute), comparative or superlative form or an *-ing* participle or *-ed* participle form.

### Examples

<i>a glimpse of the <u>obvious</u></i>	NADJ(pos)
<i>your <u>spare</u></i>	NADJ(pos)
<i>the <u>moneyed</u></i>	NADJ(pos)
<i>the <u>more advanced</u> of these</i>	NADJ(pos,edp)
<i>the <u>willing</u></i>	NADJ(pos,ingp)
<i>the <u>murdered</u></i>	NADJ(pos,edp)
<i>the <u>easier</u> of the two</i>	NADJ(comp)
<i>my <u>best</u></i>	NADJ(sup)
<i>its <u>strongest</u></i>	NADJ(sup)
<i>the <u>hungriest</u></i>	NADJ(sup)

### Notes

- when the head of the noun phrase is realised by a noun, the set of adverbs which can take the function of noun phrase premodifier is limited to a small set. However, in the case of a nominal adjective as the head of the noun phrase, a great variety of adverbs can take the function of premodifier, especially intensifying adverbs. Some examples are:

*the most favoured*

*the very worst*  
*the extremely poor*  
*the undoubtedly genuine*

- a noun can generally take various types of determiner, whereas a nominal adjective can only take a definite article, a possessive pronoun, a demonstrative pronoun, a genitive NP or a quantifying pronoun. For example,

*the rich*  
*their most favoured*  
*these rich*  
*the country's rich*

- *-ing* and *-ed* forms of nominal adjectives that are introduced by the negative prefix *un-* or *in-* are not tagged with 'ingp' or 'edp':

*the unwilling* NADJ(pos)

## 2.8 Numeral

Numerals receive the wordclass label NUM. Additional feature information relates to subclass (cardinal, ordinal, fractional, hyphenated or multiplicative) and form (singular or plural). The following tags apply:

NUM(card,sing)	NUM(frac,sing)	NUM(mult)
NUM(card,plu)	NUM(frac,plu)	
NUM(ord,sing)	NUM(hyph,sing)	
NUM(ord,plu)	NUM(hyph,plu)	

### Class characteristics

Numerals can function as postdeterminer or as head of the noun phrase.

### Examples

<i>So far, <u>seventy</u> people have called us.</i>	NUM(card,sing)
<i>He's <u>thirty-five</u></i>	NUM(card,sing)
<i>She's the <u>first</u>.</i>	NUM(ord,sing)
<i>The governments of <u>35-40</u> low income countries met in London yesterday.</i>	NUM(hyph,sing)
<i><u>Two-thirds</u> of the material had already been spoiled.</i>	NUM(frac,plu)
<i>They only met <u>twice</u> during that period.</i>	NUM(mult)

### Notes

- If the numeral has a singular *form*, it should be tagged as singular, even if its meaning is 'more than one'. For example:

<i>one</i>	NUM(card,sing)
<i>two</i>	NUM(card,sing)
<i>fifty-six</i>	NUM(card,sing)
<i>first</i>	NUM(ord,sing)
<i>second</i>	NUM(ord,sing)

Only if the numeral has a plural form is it tagged as plural. For example,

<i>ones</i>	NUM(card,plu)
-------------	---------------

<i>1920s</i>	NUM(card,plu)
<i>eighths</i>	NUM(ord,plu)
<i>quarters</i>	NUM(frac,plu)
<i>1950-60s</i>	NUM(hyph,plu)

### 2.8.1 Cardinal numeral

The subclass of cardinal numerals is a closed, but non-enumerable class. Its members include the following:

*1, 2, 3, 4, 5, ...*  
*one, two, three, four, five, ...*  
*hundred, thousand, million, ...*  
*hundreds, thousands, millions, ...*  
*zillion*  
 $\infty$

#### Examples

<i>She ate <u>three</u> cakes.</i>	NUM(card,sing)
<i><u>2</u> plus <u>2</u> equals <u>4</u>.</i>	NUM(card,sing)
<i>The man is <u>forty-four</u></i>	NUM(card,sing)
<i>1970s</i>	NUM(card,plu)
<i>'80s</i>	NUM(card,plu)

#### Note

- Cardinal numerals that are separated by means of a slash (/) should be treated as separate tokens; the slash should be tagged as a co-ordinator. For example,

*Approximately in 1995/1996.* NUM(card,sing) CONJUNC(coord) NUM(card,sing)

### 2.8.2 Ordinal numeral

The subclass of ordinal numerals is closed but non-enumerable. Its members include the following:

*1st, 2nd, 3rd, 4th, ...*  
*first, second, third, fourth, ...*  
*hundredth*

Also:

*nth, umpteenth.*

#### Class characteristics

Ordinal numerals are

1. a figure, followed by the suffixes *-st* (for numbers ending in 1), *-nd* (for numbers ending in 2), *-rd* (for numbers ending in 3) or *-th* (for number ending in 4, 5, 6, 7, 8, 9 or 0).
2. their alphabetical equivalents: *first, second, third, fourth, fifth, sixth*, etc.

#### Examples

<i>The <u>first</u> time you came here, you were wearing that skirt.</i>	NUM(ord,sing)
<i>The <u>3rd</u> bottle was empty.</i>	NUM(ord,sing)



### 2.8.3 Fractional numeral

The subclass of fractional numerals is a closed, but non-enumerable class. Its members are all fractions in numbers and words, as well as *half*, *halves*, *quarter* and *quarters*.

#### Examples

<i>2/3</i>	NUM(frac,sing)
<i>3/5s</i>	NUM(frac,plu)
<i>two-thirds</i>	NUM(frac,plu)
<i>195/250</i>	NUM(frac,sing)
<i>two and a half</i>	NUM(card,sing) CONJUNC(coord) ART(indef)
	NUM(frac,sing)

#### Notes

- A fractional numeral is generally preceded by cardinal numeral, an ordinal number by an article. Compare:

<i>one third</i>	NUM(card,sing) NUM(frac,sing)
<i>the third</i>	ART(def) NUM(ord,sing)

- In hyphenated numerals such as *1980-1885*, the meaning of the hyphen is 'up to'. Such numerals should not be confused with hyphenated fractional numerals. Compare:

<i>1980-1885</i>	NUM(hyph,sing)
<i>a <u>one-hundredth</u> part</i>	NUM(frac,sing)

### 2.8.4 Hyphenated numeral

The subclass of hyphenated numerals is an open class. A hyphenated numeral consists of two figures or numbers which are connected by means of a hyphen.

#### Examples

<i>1980-1985</i>	NUM(hyph,sing)
<i>1960-70s</i>	NUM(hyph,plu)

#### Note

- Compound numerals where the hyphen means *and* should not be tagged as hyphenated numerals, nor should hyphenated fractional numerals. Compare:

<i>1980-1985</i>	NUM(hyph,sing)
<i>fifty-five</i>	NUM(card,sing)
<i>two-thirds</i>	NUM(frac,plu)

### 2.8.5 Multiplicative

The subclass of multiplicatives is a closed class. Its members are *once*, *twice*, and *thrice* (in the sense of 'one time', 'two times' and 'three times').

#### Examples

<i>He wanted <u>twice</u> as much.</i>	NUM(mult)
<i>She came only <u>once</u>.</i>	NUM(mult)
<i>I've been there <u>thrice</u>.</i>	NUM(mult)
<i>He received <u>twice</u> the amount.</i>	NUM(mult)

**Note**

- *Once* can also be a subordinating conjunction, or an adverb, for example:

Once you've seen her, you'll understand.      CONJUNC(subord)  
Once upon a time ...                              ADV(ge,pos)

**2.9 Preposition**

Prepositions receive the major wordclass label PREP. Additional feature information relates to subclass (general or phrasal). The following tags apply:

PREP(ge)  
 PREP(phras)

**Class characteristics**

Syntactically, prepositions can be characterised as follows:

1. a preposition forms the first element of a prepositional phrase
2. it is complemented by a prepositional complement which is realised by a noun phrase, a prepositional phrase, a nominal *wh*-clause or a nominal *-ing* clause (cf. CGEL, 9.1).

**2.9.1 General preposition**

The subclass of general prepositions is a closed class. Its members are:

-	<i>circa (c.)</i>	<i>plus</i>
+	<i>cum</i>	<i>pro</i>
x	<i>despite</i>	<i>qua</i>
<i>aboard</i>	<i>down</i>	<i>re (CGEL 9.57)</i>
<i>about</i>	<i>during</i>	<i>round</i>
<i>above</i>	<i>ere</i>	<i>sans</i>
<i>across</i>	<i>except</i>	<i>save</i>
<i>afore</i>	<i>for</i>	<i>since</i>
<i>after</i>	<i>from</i>	<i>than</i>
<i>against</i>	<i>in</i>	<i>through</i>
<i>along</i>	<i>including</i>	<i>throughout</i>
<i>alongside</i>	<i>inside</i>	<i>thru</i>
<i>amid(st)</i>	<i>into</i>	<i>till</i>
<i>among(st)</i>	<i>less</i>	<i>times</i>
<i>around</i>	<i>like</i>	<i>to</i>
<i>as</i>	<i>minus</i>	<i>toward</i>
<i>astride</i>	<i>near</i>	<i>towards</i>
<i>at</i>	<i>nearby</i>	<i>'tween</i>
<i>atop</i>	<i>'neath</i>	<i>'twixt</i>
<i>bar</i>	<i>notwithstanding</i>	<i>under</i>
<i>before</i>	<i>of</i>	<i>underneath</i>
<i>behind</i>	<i>off</i>	<i>unlike</i>
<i>below</i>	<i>on</i>	<i>until</i>
<i>beneath</i>	<i>onto</i>	<i>unto</i>
<i>beside</i>	<i>opposite</i>	<i>up</i>
<i>besides</i>	<i>out</i>	<i>upon</i>
<i>between</i>	<i>outside</i>	<i>versus (v/vs.)</i>
<i>betwixt</i>	<i>over</i>	<i>via</i>
<i>beyond</i>	<i>pace</i>	<i>vis-à-vis</i>
<i>but</i>	<i>past</i>	<i>with</i>
<i>by</i>	<i>per</i>	

*within**without*

Multi-token units (cf. complex prepositions CGEL 9.10-13):

<i>according to</i>	<i>in behalf of</i> (AmE)	<i>in terms of</i>
<i>à la</i>	<i>in case of</i>	<i>in view of</i>
<i>apropos of</i>	<i>in face of</i>	<i>instead of</i>
<i>as for</i>	<i>in front of</i>	<i>irrespective of</i>
<i>as from</i>	<i>in lieu of</i>	<i>(by) means of</i>
<i>as of</i>	<i>in light of</i>	<i>on account of</i>
<i>as regards</i>	<i>in line with</i>	<i>on board</i>
<i>as to</i>	<i>in memoriam</i>	<i>on behalf of</i>
<i>because of</i>	<i>in place of</i>	<i>on pain of</i>
<i>by dint of</i>	<i>in reference to</i>	<i>on top of</i>
<i>by means of</i>	<i>in regard to</i>	<i>out of</i>
<i>by virtue of</i>	<i>in relation to</i>	<i>thanks to</i>
<i>(by) way of</i>	<i>in respect of</i>	<i>with reference to</i>
<i>care of (c/o)</i>	<i>in respect to</i>	<i>with regard to</i>
<i>ex libris</i>	<i>(in) spite of</i>	<i>with respect to</i>

**Examples**

<i>Before</i> five o'clock I want an answer.	PREP(ge)
The girl stood <i>behind</i> her father.	PREP(ge)
This is <i>below</i> every reasonable standard.	PREP(ge)
This is all <i>beyond</i> my control.	PREP(ge)
It all succeeded <i>thanks to</i> her expertise.	PREP(ge):1/2 PREP(ge):2/2

**Notes**

- As opposed to phrasal prepositions (section 2.9.2), general prepositions do not have a collocational link with the verb.

<i>He looked <u>at</u> the girl.</i>	PREP(phras)
<i>He walked <u>up</u> the stairs.</i>	PREP(ge)

- The prepositional complement generally follows the preposition, unless this is stranded.

*It was the building on the left they walked into.*

- If *on board* means *aboard*, and it is followed by a prepositional complement, it should be tagged as a multi-token complex preposition, as in

*The captain was on board the ship.* PREP(ge):1/2 PREP(ge):2/2

- Times* is a preposition in a context like:

*Three times six equals eighteen.* PREP(ge)

- Including* and *like* are classified as connective adverbs if they introduce an appositive, e.g.

<i>They bought several books, <u>like</u> Hard Times.</i>	ADV(connec)
<i>They bought several books, <u>including</u> Hard Times.</i>	ADV(connec)

- Plus* can also be a co-ordinating conjunction as in:

*You can go to London, plus you can do me a favour.* CONJUNC(coord)



## 2.9.2 Phrasal preposition

The subclass of phrasal prepositions is a closed class (cf. CGEL 16.5). Among its members are:

<i>aboard</i>	<i>beside</i>	<i>out (AmE)</i>
<i>about</i>	<i>besides</i>	<i>over</i>
<i>above</i>	<i>between</i>	<i>past</i>
<i>across</i>	<i>by</i>	<i>round</i>
<i>after</i>	<i>down</i>	<i>through</i>
<i>along</i>	<i>for</i>	<i>to</i>
<i>alongside</i>	<i>from</i>	<i>toward(s)</i>
<i>against</i>	<i>in</i>	<i>under</i>
<i>among</i>	<i>into</i>	<i>up</i>
<i>around</i>	<i>like</i>	<i>upon</i>
<i>as</i>	<i>of</i>	<i>with</i>
<i>at</i>	<i>off</i>	<i>without</i>
<i>before</i>	<i>on</i>	
<i>behind</i>	<i>onto</i>	

### Class characteristics

Phrasal prepositions have a collocational and semantic link with a verb.

### Examples

<i>He looked <u>at</u> it.</i>	PREP(phras)
<i>She fell <u>for</u> him.</i>	PREP(phras)

### Notes

- Phrasal prepositions and phrasal adverbs:

1. The major difference between phrasal prepositions and phrasal adverbs is brought out by pronominalization of the object NP. Consider the following examples:

<i>The student <u>looked up</u> the word in the dictionary.</i>	VB(lex,montr,past) ADV(phras)
<i>The student <u>looked at</u> the word in the dictionary.</i>	VB(lex,montr,past) PREP(phras)

Pronominalization of the object NP results in:

<i>The student looked it <u>up</u> in the dictionary.</i>	ADV(phras)
<i>*The student looked <u>up</u> it in the dictionary.</i>	
<i>The student looked <u>at</u> it in the dictionary.</i>	PREP(phras)
<i>*The student looked it <u>at</u> in the dictionary.</i>	

2. A phrasal adverb does not form one constituent with a following NP, e.g.

<i>The student <u>looked up</u> the word in the dictionary</i>	VB(lex,montr,past) ADV(phras)
--	-------------------------------

where the cleft form is impossible:

*\*It was up the word in the dictionary that the student looked.*

3. A phrasal preposition constitutes a PP with the following NP, e.g.

<i>The student <u>looked at</u> the word in the dictionary.</i>	VB(lex,intr,past) PREP(phras)
---	-------------------------------

where a cleft form with fronted a PP is possible:

*It was at the word in the dictionary that the student looked.*

## 2.10 Proform

Proforms receive the wordclass label PROFM. Additional feature information relates to subclass (conjoin, one, so). Only for the subclass 'one' a second feature indicating its form (singular, plural) is used. The subclass 'so' receives a second feature indicating the nature of the substitution (phrase, clause). The following tags apply:

PROFM(conj)  
 PROFM(one,sing)  
 PROFM(one,plu)  
 PROFM(so,phrase)  
 PROFM(so,clause)

### Class characteristics

Proforms are substitutes for phrases or clauses. After a coordinator, they replace a conjoin. The proform *one(s)* is a substitute for the head of a noun phrase.

### 2.10.1 Pro-conjoin

The subclass of pro-conjoins is a closed class. Its members are:

<i>all</i>	<i>not</i>	<i>whatever</i>
<i>both</i>	<i>otherwise</i>	<i>whichever</i>
<i>etc.</i>	<i>something</i>	
<i>neither</i>	<i>suchlike</i>	

multi-token units:

<i>et al.</i>	<i>so forth</i>	<i>the reverse</i>
<i>et cetera</i>	<i>such like</i>	<i>vice versa</i>
<i>so on</i>	<i>the like</i>	<i>what have you</i>

### Class characteristics

Members from this class always follow an explicit coordinator (*and*, *but* or *or*), with the exception of *etc.*, *et cetera* and *et al.*

### Examples

*They sold spiders, snakes and so on.*  
*The agreement imposes restrictions on prices to be charged, quantities to be produced and so forth.*  
*It is not possible to deduce structures from functions and vice versa.*  
*Whether that small amount was enough to judge the test conclusively or not, was unclear.*  
*It wouldn't be him leaving her but the reverse.*  
*You deaf, stupid, or both?*

### 2.10.2 Proform one

The subclass of proform *one* is a closed class. Its members are: *one*, *ones*, *'un*, and *'uns*.

### Class characteristics

Members from this class substitute for the head of the noun phrase. *One* and *'un* are singular, *ones* and *'uns* are plural.

### Examples

<i>That other <u>one</u>'s being boiling for half an hour.</i>	PROFM(one,sing)
<i>This is the only way of saving our way of life and the very lives of our loved <u>ones</u>.</i>	PROFM(one,plu)

### Note

- *one* can also be a cardinal numeral: NUM(card,sing), or a pronoun PRON(one).

<i>Have you seen the blue <u>one</u>?</i>	PROFM(one,sing)
<i>I saw <u>one</u> dolphin.</i>	NUM(card,sing)
<i><u>One</u> cannot deal with such problems on one's own.</i>	PRON(one)

### 2.10.3 Proform so

The subclass of proform *so* is closed. Its members are *so* and *not*.

### Class characteristics

The proform *so* is used to replace phrases or clauses, and receives either of the corresponding features 'phrase' or 'clause'. The form *not* can only replace clauses.

### Examples

<i>The man's wife had told Judith that he was tense and jumpy, and <u>so</u> he was.</i>	PROFM(so,phrase)
<i>The boys promised to be good and <u>so</u> they were.</i>	PROFM(so,phrase)
<i>This proved to be <u>so</u>.</i>	PROFM(so,phrase)
<i>I don't think <u>so</u>, not in your case.</i>	PROFM(so,clause)
<i>I've always assumed <u>so</u>, although perhaps on a rather narrow basis of experience.</i>	PROFM(so,clause)
<i>Do you think there is going to be a strike? I believe <u>so</u>.</i>	PROFM(so,clause)
<i>Just why he did <u>so</u>, he didn't clearly know.</i>	PROFM(so,phrase)
<i>But I'd suppose <u>not</u>.</i>	PROFM(so,clause)
<i>The chap thinks definitely <u>not</u>.</i>	PROFM(so,clause)

### Note

- *so* can also be a general adverb, a subordinating conjunction or a connective adverb. Examples are:

<i>I'm <u>so</u> tired.</i>	ADV(ge,pos)
<i>They've arrived, <u>so</u> you can open a bottle of wine now.</i>	CONJUNC(subord)
<i><u>So</u>, who do you think you are!</i>	ADV(connec)

### 2.11 Pronoun

Pronouns receive the major wordclass label PRON. Additional feature information relates to subclass (anticipatory *it*, cleft *it*, assertive, demonstrative, exclamatory, interrogative, negative, non-assertive, nominal possessive, *one*, personal, possessive, quantifying, reciprocal, relative, *-self*, *such*, or universal). It is with the subclasses of demonstrative, personal, possessive, nominal possessive and *-self* pronouns that a feature is assigned that indicates singular or plural form. When the form is the same for singular and plural, the feature 'number' is assigned. With anticipatory *it*, cleft *it*, and the singular personal pronoun *it*

the feature 'proclitic' can be found, while it is only with plural personal pronouns that the feature 'enclitic' can be assigned. The following tags apply:

PRON(ass)	PRON(such)	PRON(nomposs,sing)
PRON(exclam)	PRON(univ)	PRON(nomposs,plu)
PRON(inter)	PRON(antit)	PRON(nomposs,number)
PRON(inter,poss)	PRON(antit,procl)	PRON(self,sing)
PRON(neg)	PRON(cleft)	PRON(self,plu)
PRON(nonass)	PRON(cleft,procl)	PRON(pers,sing)
PRON(one)	PRON(dem,number)	PRON(pers,sing,procl)
PRON(quant)	PRON(dem,sing)	PRON(pers,plu)
PRON(recip)	PRON(dem,plu)	PRON(pers,plu,encl)
PRON(rel)	PRON(poss,sing)	PRON(pers,number)
PRON(rel,poss)	PRON(poss,plu)	

### Class characteristics

Pronouns, like other proforms, are words that fill slots in clauses or phrases without carrying any individual semantic content. Sometimes their only function is to fill the slot, but usually they acquire their meaning in the sentence through verbal or situational context. Syntactically, they can function as head of a noun phrase or as determiner.

### Notes

- The feature 'proclitic' is assigned to the pronoun *it*, when there is elision of the *i* ('t) and the form is contracted with the following word. The contracted form is tagged as comprising two uncontracted forms.

*And 'tis very nice indeed.* PRON(pers,sing,procl)  
*'Twas early in May.* PRON(pers,sing,procl)

- The feature 'enclitic' is assigned to enclitic forms of the personal pronouns '*em* (*them*) and '*s* (*us*):

*Stick'em up!* PRON(pers,plu,encl)  
*Let's go!* PRON(pers,plu,encl)

- The features 'singular' and 'plural' are assigned to all pronouns that distinguish in form between singular and plural. 'number' is assigned to pronouns that have the same form for singular and plural. For example,

*That's my business.* PRON(dem,sing)  
*Can you show me the way?* PRON(pers,number)  
*The judgment of sick people is sometimes not like ours.* PRON(nomposs,plu)

- The form *whose* can be an interrogative or a relative pronoun. In both cases it is tagged with the feature 'possessive'. For example,

*Whose car is that?* PRON(inter,poss)  
*The man whose wife was killed in the crash has moved away.* PRON(rel,poss)

### 2.11.1 Anticipatory *it*

The subclass of anticipatory *it* is a closed class. Its only member is *it*.



### Class characteristics

The (antit) tag is only assigned in sentences or clauses with an extraposed subject or object clause.

### Examples

<i><u>It</u> would be a shade awkward for the police to know about it.</i>	PRON(antit)
<i><u>It</u> was up to him to deliver them the goods.</i>	PRON(antit)
<i>Appleby thought <u>it</u> unnecessary to enter upon it now.</i>	PRON(antit)
<i>She no doubt judged <u>it</u> odd that he stayed away.</i>	PRON(antit)
<i><u>'T</u>was on a dark night that Graham left, never to be seen again.</i>	PRON(antit,procl)
<i>And <u>'t</u>is very nice indeed to be here.</i>	PRON(antit,procl)

### 2.11.2 Assertive pronoun

The subclass of assertive pronouns is a closed class. Its members are:

<i>some</i>	<i>someone</i>	<i>summat</i>
<i>somebody</i>	<i>something</i>	

There are no multi-token units.

### Class characteristics

An assertive pronoun either functions as a determiner (*some, certain*) or as the head of a simple noun phrase (*some, somebody, someone, something/summat*).

### Examples

<i><u>Somebody</u> was out there.</i>	PRON(ass)
<i>You must do <u>something</u> about it.</i>	PRON(ass)

### 2.11.3 Cleft *it*

This subclass is closed. Its only member is *it*.

### Class characteristics

The (cleft) tag is only assigned to *it* in so-called cleft sentences, in which one particular constituent is emphasized by putting it in the position after *it + be*, so that the resulting sentence or clause is of the pattern: *it + be + emphasized constituent + who/that ...*

### Examples

<i><u>It</u> was a child's face that looked at him now.</i>	PRON(cleft)
<i><u>'T</u>was the mailman that was responsible.</i>	PRON(cleft,procl)

### Notes

- The verb *be* in a cleft sentence or clause is tagged as intransitive:

<i>It <u>was</u> Hitler's war that did it.</i>	VB(lex,intr,past)
<i>It would <u>be</u> Mrs Martineau who would be chiefly horrified.</i>	VB(lex,intr,infin)

- It* can also be a personal pronoun or anticipatory *it* (see sections 2.11.11 and 2.11.1 resp.)

<i>I saw <u>it</u> yesterday</i>	PRON(pers,sing)
<i><u>It</u> was argued that he had failed her.</i>	PRON(antit)

### 2.11.4 Demonstrative pronoun

The subclass of demonstrative pronouns is a closed class. Its members are:

<i>that</i>	<i>these</i>	<i>those</i>
<i>them</i>	<i>this</i>	

and the archaic forms *yon* and *yonder* (both receive the tag PRON(dem,number))

There are no multi-unit tokens.

#### Class characteristics

Demonstrative pronouns function as determiners or as heads of simple noun phrases.

#### Examples

<u><i>This</i></u> is the best I've ever heard.	PRON(dem,sing)
<u><i>Those</i></u> people I've never met.	PRON(dem,plu)

#### Notes

- Generally speaking, *them* is a personal pronoun, as in *He saw them*. However, it can function as a determiner, in which case *them* is a demonstrative pronoun, as in *He saw them boys*.

### 2.11.5 Exclamatory pronoun

The subclass of exclamatory pronouns is a closed class. Its member is *what*.

#### Class characteristics

The exclamatory pronoun tag is assigned to *what* used in exclamatory phrases or sentences.

#### Examples

<u><i>What</i></u> a wonderful solution they have found!	PRON(exclam)
<u><i>What</i></u> a shame.	PRON(exclam)

### 2.11.6 Interrogative pronoun

The subclass of interrogative pronouns is a closed class. Its members are:

<i>what</i>	<i>whichever</i>	<i>whom</i>
<i>whatever</i>	<i>who</i>	<i>whomever</i>
<i>which</i>	<i>whoever</i>	<i>whose</i>

There are no multi-token units.

#### Class characteristics

This tag is assigned to *wh*-words in dependent and independent questions. They function as determiner or as head of a simple noun phrase.

**Examples**

<i><u>Which</u> train are you taking?</i>	PRON(inter)
<i><u>What</u> are you doing?</i>	PRON(inter)
<i>He asked me <u>what</u> I was doing.</i>	PRON(inter)
<i>He asked me <u>what</u> was I doing in the basement.</i>	PRON(inter)
<i>I wonder <u>who</u>'ll turn up.</i>	PRON(inter)
<i>The question is <u>who</u> we should invite.</i>	PRON(inter)

**Note**

- It is not always easy to distinguish between a *wh*-word as an interrogative in a 'dependent question' and as a nominal relative (see section 2.11.15). As the examples show, the dependent interrogative pronoun occurs basically in two environments:
  - at the beginning of an object clause after a reporting verb (*ask*, *wonder*, etc.);
  - at the beginning of a subject complement clause, after a (subject) noun phrase with a head like *question*, *problem*, *mystery*, etc.
- Whose* is a possessive interrogative pronoun and should be tagged PRON(inter,poss).

**2.11.7 Negative pronoun**

The subclass of negative pronouns is a closed class. Its members are:

<i>neither</i>	<i>none</i>
<i>no</i>	<i>nothing</i>
<i>nobody</i>	

and the multi-token unit *no one*

**Class characteristics**

The negative pronouns all have a negative meaning component. *No* only occurs as a determiner, *none*, *nobody*, *nothing* and *no one* only as heads of NPs, while *neither* can function both as determiner and as head of an NP.

**Examples**

<i>I mean <u>no</u> more than I say.</i>	PRON(neg)
<i>There was <u>nobody</u> there.</i>	PRON(neg)
<i>There was <u>nothing</u> impulsive about Martine.</i>	PRON(neg)
<i><u>Neither</u> of us knew.</i>	PRON(neg)

**2.11.8 Nominal possessive pronoun**

The subclass of nominal possessive pronouns is a closed class. Its members are:

<i>mine</i>	<i>hers</i>	<i>theirs</i>
<i>yours</i>	<i>his</i>	
<i>thine</i>	<i>ours</i>	

There are no multi-token units.

**Class characteristics**

The nominal possessives can only be used as heads of noun phrases.

**Examples**

<i><u>Mine</u> doesn't do things like that.</i>	PRON(nomposs,sing)
<i>That was <u>ours</u> to begin with.</i>	PRON(nomposs,plu)
<i>This book is <u>yours</u>.</i>	PRON(nomposs,number)

**Note**

- Both *yours* and *thine* are tagged PRON(nomposs,number).

**2.11.9 Non-assertive pronoun**

The subclass of non-assertive pronouns is a closed class. Its members are:

<i>any</i>	<i>anyone</i>	<i>either</i>
<i>anybody</i>	<i>anything</i>	

There are no multi-token units.

**Class characteristics**

A non-assertive pronoun either functions as a determiner (*any*, *either*) or as the head of a simple noun phrase (*any*, *anybody*, *anyone*, *anything*, *either*).

**Examples**

<i>It's impossible to see <u>anything</u> from the inside.</i>	PRON(nonass)
<i>Simona does her best to monopolize <u>any</u> young men there are.</i>	PRON(nonass)
<i>Can <u>either</u> of you lend me a hand?</i>	PRON(nonass)

**Note**

- *any* can also be an adverb. For example,

<i>I really can't stay <u>any</u> longer.</i>	ADV(ge,pos)
---	-------------

**2.11.10 Pronoun one**

The subclass of pronoun *one* is closed. Its only member is *one*.

**Class characteristics**

The (pronoun *one*) tag is only assigned to occurrences of *one* where it means 'people' in general. This implies that it never takes the plural form *ones*. It always constitutes a one-word noun phrase.

**Examples**

<i><u>One</u> doesn't do things like that.</i>	PRON(one)
<i>What <u>one</u> could do under such circumstances is quite a lot.</i>	PRON(one)

**2.11.11 Personal pronoun**

The subclass of personal pronouns is a closed class. Its members are:

<i>I</i>	<i>she</i>	<i>him</i>	<i>you</i>
<i>me</i>	<i>s/he</i>	<i>we</i>	<i>they</i>
<i>you</i>	<i>it</i>	<i>us</i>	<i>them</i>
<i>he</i>	<i>her</i>	<i>'s</i>	<i>'em</i>

and the archaic forms: *thou*, *thee*, and *ye*.

There are no multi-token units.

### Class characteristics

The personal pronouns occur as heads of simple NPs which usually contain just the head. They are marked for person (1st, 2nd, and 3rd) and also (with the exception of *you* and *it*) for case (subjective and objective case) and number (singular and plural). The third person singular pronouns are also marked for gender (masculine, feminine and neuter).

### Examples

<i><u>He</u> became alarmed.</i>	PRON(pers,sing)
<i>Usually, <u>they</u> played piquet.</i>	PRON(pers,plu)
<i>I am sure <u>you</u> will have noticed it.</i>	PRON(pers,number)

### Notes

- 'formal' *it* is treated as a regular personal pronoun. For example,

<i><u>It</u> was raining.</i>	PRON(pers,sing)
<i><u>It</u> was rather cold outside.</i>	PRON(pers,sing)
<i><u>It's</u> a quarter past three.</i>	PRON(pers,sing)
<i><u>It's</u> morning already.</i>	PRON(pers,sing)

- *them* can also be a demonstrative pronoun, as in:

<i>He met <u>them</u> boys in Birmingham.</i>	PRON(dem,plu)
---	---------------

- Although the slash generally is a coordinator, the following form receives only one tag:

<i>s/he</i>	PRON(pers,sing)
-------------	-----------------

- See also: possessive pronoun, formal *it*, cleft *it*, and anticipatory *it*.

### 2.11.12 Possessive pronoun

The subclass of possessive pronouns is a closed class. Its members are:

<i>my</i>	<i>his</i>	<i>your</i>
<i>your</i>	<i>its</i>	<i>their</i>
<i>her</i>	<i>our</i>	

and the archaic form *thy*.

There are no multi-token units.

### Class characteristics

Like personal pronouns, the possessive pronouns also show distinctions for person, number and gender. They function as determiners in NPs.

### Examples

<i>They told him <u>his</u> wife had been drowned.</i>	PRON(poss,sing)
<i><u>Their</u> fingerprints were found on the handle-bar.</i>	PRON(poss,plu)

### 2.11.13 Quantifying pronoun

The subclass of quantifying pronouns is a closed class. Its members are:

<i>enough</i>	<i>little</i>	<i>most</i>
<i>few</i>	<i>least</i>	<i>plenty</i>
<i>fewer</i>	<i>much</i>	<i>several</i>
<i>fewest</i>	<i>many</i>	
<i>less</i>	<i>more</i>	

There are no multi-token units.

#### Class characteristics

The quantifying pronouns are semantically distinguished from other pronouns in denoting number or quantity.

#### Examples

*He had never carried much weight physically and was unarmed.* PRON(quant)  
*The younger you are the fewer facts of which you are aware.* PRON(quant)

### 2.11.14 Reciprocal pronoun

The subclass of reciprocal pronouns is a closed class. Its members are the multi-token units *each other* and *one another*.

#### Class characteristics

The reciprocal pronouns occur as heads of one-word noun phrases in non-subject functions. They are used in sentences with plural or coordinated subjects.

#### Examples

*They know each other quite well.* PRON(recip):1/2 PRON(recip):2/2  
*They hadn't seen one another for years.* PRON(recip):1/2 PRON(recip):2/2

### 2.11.15 Relative pronoun

The subclass of relative pronouns is a closed class. Its members are:

#### with antecedent:

<i>that</i>	<i>who</i>	<i>whose</i>
<i>which</i>	<i>whom</i>	

#### without antecedent:

<i>what</i>	<i>which</i>	<i>whoever</i>
<i>whatever</i>	<i>whichever</i>	<i>whosoever</i>
<i>whatsoever</i>	<i>whichever</i>	

There are no multi-token units.

#### Class characteristics

The class of relative pronouns comprises two subclasses, which, however, do not receive distinct tags. Members of the first subclass occur with an antecedent and introduce relative clauses functioning as postmodifiers in noun phrases. The second subclass ( the 'nominal

relatives') occurs without an antecedent and introduces clauses that have independent sentence functions.

### Examples

<i>He discovers that it is Fell <u>who</u> is the local medical practitioner.</i>	PRON(rel)
<i>Let me tell you of something <u>that</u> Friary said.</i>	PRON(rel)
<i>There was an awkward moment in <u>which</u> conversation failed to happen.</i>	PRON(rel)
<i><u>What</u> chiefly struck Appleby now was something quite different.</i>	PRON(rel)

### Notes

- *Whose* is a possessive relative pronoun and is tagged as PRON(rel,poss).

### 2.11.16 -self pronoun

The subclass of -self pronouns is a closed class. Its members are:

<i>myself</i>	<i>himself</i>	<i>yourselves</i>
<i>yourself</i>	<i>itself</i>	<i>themselves</i>
<i>herself</i>	<i>ourselves</i>	<i>oneself</i>

and the archaic form *thysself*.

There are no multi-token units.

### Class characteristics

The -self pronouns show distinctions for person, number and gender, but not for case. The -self pronoun corresponding to the pronoun *one* is *oneself*. They can be used emphatically as well as non-emphatically. In the latter case they usually function as independent sentence constituents. All -self pronouns end in the singular suffix -self or the plural suffix -selves.

### Examples

<i>She watched <u>herself</u> in the mirror..</i>	PRON(self,sing)
<i>I made a painting of <u>myself</u>.</i>	PRON(self,sing)
<i>You shouldn't hurt <u>yourself</u>.</i>	PRON(self,sing)
<i>He did it <u>himself</u>.</i>	PRON(self,sing)
<i>I <u>myself</u> don't like it at all.</i>	PRON(self,sing)
<i>They wouldn't touch it <u>themselves</u>.</i>	PRON(self,plu)

### 2.11.17 Pronoun such

The subclass of pronoun *such* is closed. Its only member is *such*.

### Class characteristics

*Such* can occur as determiner in the noun phrase. In one-word noun phrases it functions as head.

### Examples

<i><u>Such</u> people have lived here for ages.</i>	PRON(such)
<i>I wouldn't mind <u>such</u> a car.</i>	PRON(such)
<i><u>Such</u> filters are extremely unreliable.</i>	PRON(such)
<i><u>Such</u> was his indignation that he slammed down the telephone.</i>	PRON(such)

### 2.11.18 Universal pronoun

The subclass of universal pronouns is a closed class. Its members are:

*all*  
*both*  
*each*

*every*  
*everybody*  
*everyone*

*everything*

There are no multi-token units.

### Class characteristics

Universal pronouns function as determiners in noun phrases. They also occur as heads in one-word noun phrases.

### Examples

All this was being thrown at him. PRON(univ)  
Everything around was quite still. PRON(univ)

## 2.12 Particle

Particles receive the major wordclass label PRTCL. Additional feature information relates to subclass (for, to, with). The following tags apply:

PRTCL(for)  
PRTCL(to)  
PRTCL(with)

### 2.12.1 Particle for

The subclass of particles *for* is a closed class. Its members are *for* and *in order for*.

### Class characteristics

The particles *for* and *in order for* introduce non-finite (infinitive) clauses with explicit subjects.

### Examples

*I put out my arms for her to come to me.* PRTCL(for)  
*As he waited for the meeting to begin,*  
*Holman looked down the length of the long oak table.* PRTCL(for)  
*I waited for the girl to call me.* PRTCL(for)

### 2.12.2 Particle to

The subclass of particles *to* is a closed class. Its members are *to*, *in order to* and *so as to*.

### Class characteristics

The particles *to*, *in order to* and *so as to* introduce non-finite (infinitive) clauses.

### Examples

*She told me to go away.* PRTCL(to)  
*He shouted so as to draw my attention.*  
PRTCL(to):1/3 PRTCL(to):2/3 PRTCL(to):3/3  
*He turned around so as to see who was coming in.*  
PRTCL(to):1/3 PRTCL(to):2/3 PRTCL(to):3/3  
*They called in order to hear what they should do about it.*  
PRTCL(to):1/3 PRTCL(to):2/3 PRTCL(to):3/3



### 2.12.3 Particle *with*

The subclass of particles *with* is a closed class. Its members are *with*, *without*, *what with* and *what without*.

#### Class characteristics

The particles *with*, *without*, *what with* and *what without* introduce non-finite clauses or verbless clauses.

#### Examples

<i>We can't have a party <u>with</u> a dead body lying on the doorstep.</i>	PRTCL(with)
<i>I'd prefer to brush my teeth <u>without</u> anybody watching.</i>	PRTCL(with)
<i><u>With</u> John coming home, I didn't know what to do.</i>	PRTCL(with)
<i><u>Without</u> my assistant being present, I felt pretty tense.</i>	PRTCL(with)
<i><u>What with</u> her husband in prison and her son in America, she is a bit lonely.</i>	PRTCL(with):1/2 PRTCL(with):2/2

### 2.13 Verb

Verbs receive the major wordclass label VB. Additional feature information relates to type of verb (auxiliary verb, lexical verb), subclass (for auxiliary verbs: do, modal, passive, perfective, progressive, semi, semi followed by -ing participle; for lexical verbs: intransitive, copular, monotransitive, ditransitive, dimono transitive, complex transitive) and form (infin, pres, past, ingp, edp, subjun; encl, neg, procl; ellipt). The tags that apply are given at the beginning of the subsections in which the different verb classes are described.

#### Notes

- The feature 'encl' (enclitic) is assigned to enclitic forms of verbs. An enclitic form is one which is contracted with a preceding noun or pronoun. The first part of the verb has been replaced by an apostrophe. The enclitic form of a verb is treated as a separate token. For example,

<i>We're waiting for John.</i>	VB(aux,prog,pres,encl)
<i>It'll change for the better, I'm sure.</i>	VB(aux,modal,pres,encl)
<i>You've done enough for today.</i>	VB(aux,perf,pres,encl)
<i>He's an accountant.</i>	VB(lex,cop,pres,encl)

- The feature 'neg' (negative) is assigned to auxiliaries which show negative contraction with the full form *not* (e.g. *cannot*) or the enclitic form *n't* (e.g. *isn't*). The contracted negative form is not treated as a separate token, i.e. the negative adverb is not tagged as a separate token. Instead the auxiliary is tagged with the negative feature 'neg'. For example,

<i>He <u>won't</u> be staying much longer.</i>	VB(aux,modal,pres,neg)
<i>He <u>isn't</u> coming.</i>	VB(aux,prog,pres,neg)
<i>You really <u>can't</u> do this to him.</i>	VB(aux,modal,pres,neg)
<i>It can, but then again, it <u>cannot</u>.</i>	VB(aux,modal,pres,neg)
<i>You will tell her, <u>won't</u> you?</i>	VB(aux,modal,pres,neg)
<i>They <u>aren't</u> as sweet as I thought.</i>	VB(lex,cop,pres,neg)
<i>He <u>wasn't</u> a butcher at all.</i>	VB(lex,cop,past,neg)
<i>He <u>hasn't</u> a penny on him.</i>	VB(lex,cxtr,pres,neg)

- The feature 'subjun' (subjunctive) is assigned only to verbs that are marked for subjunctive. For example,

<i>If he <u>were</u> leaving, then perhaps you could replace him.</i>	VB(aux,prog,subjun)
<i>They suggested that he <u>be</u> appointed.</i>	VB(aux,pass,subjun)
<i>He insisted that she <u>go</u>.</i>	VB(lex,intr,subjun)

The subjunctive feature is not assigned to unmarked forms even though strictly speaking these are subjunctives as well (as is shown by the fact that otherwise *should* is required). For example,

<i>I insist that you <u>go</u>.</i>	VB(lex,intr,pres)
<i>I suggest he <u>leave</u> at once.</i>	VB(lex,intr,subjun)
<i>Home is home, <u>be</u> it every so homely.</i>	VB(lex,cop,subjun)

### 2.13.1 Auxiliary verb

There are seven subclasses of auxiliary verbs. These are: *do*, modal, perfective, passive, progressive, semi and semi followed by an *-ing* participle.

#### Class characteristics

Syntactically, auxiliaries can be characterised as follows:

1. in conditional clauses operators in declarative constructions, auxiliaries admit inversion of subject and operator (cf. CGEL 3.24). For example,

<i><u>Had</u> he known this, he would never have left.</i>	VB(aux,perf,past)
--	-------------------

2. frequency subjuncts (e.g. *always*, *never*) and disjuncts (e.g. *certainly*, *probably*) must precede lexical verbs, while they may follow (the first part of) an auxiliary functioning as operator. For example,

<i>They always <u>came</u> early.</i>	VB(lex,intr,past)
<i>They <u>would</u> always come early.</i>	VB(aux,modal,past)
<i>They <u>weren't</u> ever going to be early.</i>	VB(aux,semi,past,neg):1/3

Prosodic/semantic characteristics: as operators, auxiliaries can carry nuclear stress to mark a finite clause as positive rather than negative (cf. CGEL 3.25). For example,

<i>Won't you try again? Yes, I <u>WILL</u> try again.</i>	VB(aux,modal,pres)
---	--------------------

#### Semantic characteristics:

1. auxiliaries are semantically independent of the subject, i.e. they cannot be assigned selection features, like lexical verbs. Unlike lexical verbs, auxiliaries do not have independent meaning.
2. an auxiliary can be followed by the negative adverb *not*, which is not possible for lexical verbs.

*He is not coming to Paris.*  
*\*He came not to Paris.*

However, there is an exception for *be* and for *have* (in the sense of 'to possess'): For example:

*He is not a fool.*  
*He is not!*  
*(He has a car.) No, he has not!*

Morphological characteristics: auxiliaries have contracted negative forms: *isn't*, *can't*, *isn't going to*, *won't*, etc. In addition, many auxiliaries have contracted non-negative forms: *'m*, *'ll*, *'s going to*, *'re*, *'ve*, etc. (see CGEL, 3.23). However, the lexical verbs *be* and *have* have the same characteristics. For example:

<i>He <u>isn't</u> rich.</i>	VB(lex,cop,past,neg)
<i>She <u>hasn't</u> a lot of patience.</i>	VB(lex,montr,pres,neg)

### Discontinuous auxiliary verb

If the constituent parts of a multi-token auxiliary verb do not occur adjacent to each other, the verb is discontinuous. Each token of the multi-token unit should be tagged by means of a ditto tag. When making manual correction, make sure that the numbers are correct. The first number should denote the position of the token in the unit; the second the total number of tokens in the unit. The tokens which interrupt discontinuous auxiliaries should receive their appropriate tags. For example,

*He wasn't ever going to forget what had happened.* This is tagged as follows:

<i>wasn't</i>	VB(aux,semi,past,neg,disc):1/3
<i>ever</i>	ADV(ge,pos)
<i>going</i>	VB(aux,semi,past,neg,disc):2/3
<i>to</i>	VB(aux,semi,past,neg,disc):3/3

#### 2.13.1.1 Do: auxiliary of periphrasis and emphasis

The subclass of auxiliary verb *do* is a closed class. Its only member is *do*. The following tags apply:

- VB(aux,do,imp)
- VB(aux,do,imp,neg)
- VB(aux,do,past)
- VB(aux,do,past,neg)
- VB(aux,do,pres)
- VB(aux,do,pres,encl)
- VB(aux,do,pres,neg)
- VB(aux,do,pres,procl)

#### Class characteristics

##### Syntactic characteristics:

1. the *do* auxiliary is the operator in interrogative and negative constructions as auxiliary of periphrasis; *do*-periphrasis (also: *do*-support) applies to the use of *do* as an 'empty' or 'dummy' operator (cf. CGEL section 3.37).

<i><u>Do</u> you know what her name is?</i>	VB(aux,do,pres)
<i>I <u>don't</u> know.</i>	VB(aux,do,pres,neg)

2. the auxiliary of periphrasis is used in negative sentences with *not*, in interrogative sentences (with the exception of *wh*-questions opening with the subject) and in declarative sentences opening with a negative adverbial with inversion of *do* and subject (cf. Aarts & Aarts, 1982: 37).

<i>Rarely <u>did</u> they leave the house.</i>	VB(aux,do,past)
<i><u>Don't</u> be a fool now</i>	VB(aux,do,imp,neg)

*Please, don't go!* VB(aux,do,imp,neg)  
*Don't you try it!* VB(aux,do,imp,neg)

3. *do* can be used emphatically in positive declaratives and imperatives. It can also occur when a *wh*-question opens with the subject, while its presence is not required for the grammaticality of the structure. Compare:

*I do know what her name is.* VB(aux,do,pres)  
*I know what her name is.* VB(lex,montr,pres)

*Do come in!* VB(aux,do,imp)  
*Come in!* VB(lex,intr,imp)

*Who did come in after you?* VB(aux,do,past)  
*Who came in after you?* VB(lex,intr,past)

### Examples

*He does not realise what he is doing.* VB(aux,do,pres)  
*I didn't see Geoffrey last night.* VB(aux,do,past,neg)  
*Did you know she was coming?* VB(aux,do,past)  
*Didn't you know she was coming?* VB(aux,do,past,neg)  
*Rarely did they leave the house.* VB(aux,do,past)  
*He did say that, actually.* VB(aux,do,past)  
*Do come in!* VB(aux,do,imp)  
*Do sit down.* VB(aux,do,imp)  
*Who does understand him?* VB(aux,do,pres)  
*What then did cause the explosion?* VB(aux,do,past)

### Notes

- The auxiliary *do* is sometimes contracted with the personal pronoun *you*. The final part of the auxiliary *do* is then replaced by an apostrophe, e.g.

*D'you want it?* VB(aux,do,pres,procl)  
*D'you know what I know?* VB(aux,do,pres,procl)

- do* can also be a monotransitive lexical verb which is accompanied by a direct object, as in:

*What are you doing?* VB(lex,montr,ingp)  
*Why did you do that?* VB(lex,montr,infin)  
*I didn't know what to do.* VB(lex,montr,infin)  
*I haven't done much today, really.* VB(lex,montr,edp)

- In pseudo-cleft constructions *do* is a monotransitive lexical verb (cf. CGEL 18.29):

*What he's doing is teaching him a lesson.* VB(lex,montr,ingp)

- do* can be used as an intransitive lexical verb in pro-predication (cf. CGEL 12.22). An example is:

*She didn't earn so much as she might have done.* VB(lex,intr,edp)

### 2.13.1.2 Passive auxiliary

The subclass of passive auxiliaries is a closed class. Its members are *be* and *get*. The following tags apply:

VB(aux,pass,edp)  
 VB(aux,pass,imp)  
 VB(aux,pass,infin)  
 VB(aux,pass,ingp)  
 VB(aux,pass,past)  
 VB(aux,pass,past,neg)  
 VB(aux,pass,pres)  
 VB(aux,pass,pres,encl)  
 VB(aux,pass,pres,neg)  
 VB(aux,pass,subjun)  
 VB(aux,pass,subjun,neg)

#### Examples

<i>He <u>was</u> fired.</i>	VB(aux,pass,past)
<i>He had <u>been</u> fired before.</i>	VB(aux,pass,edp)
<i>He <u>got</u> fired.</i>	VB(aux,pass,past)
<i>He failed to <u>be</u> admitted.</i>	VB(aux,pass,infin)

#### Notes

- Both *be* and *get* may also be copular lexical verbs, e.g.:

<i>He <u>got</u> drunk.</i>	VB(lex,cop,past)
<i>He <u>is</u> my best friend</i>	VB(lex,cop,past)

- Get* can also be a lexical verb, as in:

<i>He <u>got</u> a fine</i>	VB((lex,montr,past)
<i>He <u>got</u> me a glass of water.</i>	VB(lex,ditr,past)
<i>He <u>got</u> her appointed.</i>	VB(lex,cxtr,past)

- 's can be a contracted form of both *has* and *is*. 's as lexical verb is always the contracted form of *is*, while 's as auxiliary verb is either *is* or *has*.

<i>He'<u>s</u> been beaten up.</i>	VB(aux,perf,pres)
<i>He'<u>s</u> beating up someone.</i>	VB(aux,prog,pres)
<i>He'<u>s</u> alive and kicking.</i>	VB(lex,cop,pres,encl)

- Similarly, *ain't* can be both a perfective and a progressive auxiliary, or it can be a lexical verb. For example:

<i>He <u>ain't</u> going.</i>	VB(aux,prog,pres,neg)
<i>He <u>ain't</u> done that.</i>	VB(aux,perf,pres,neg)
<i>He <u>ain't</u> dead yet.</i>	VB(lex,cop,pres,neg)

- The passive auxiliary may be accompanied by a progressive auxiliary; the first *be* is the progressive auxiliary, the second the passive auxiliary, e.g.

<i>I feel I <u>am</u> <u>being</u> taken for a ride.</i>	VB(aux,prog,pres) VB(aux,pass,infin)
--	--------------------------------------

The auxiliary may form part of a reduced construction where the lexical verb is absent. In most cases the context indicates that the reduced verb phrase contains an operator realised by a passive auxiliary, as in:

*He was taken for a ride. He really was.* VB(aux,pass,past)

### 2.13.1.3 Progressive auxiliary

The subclass of progressive auxiliary verb is a closed class. Its only member is *be*. The tags that apply are the following:

VB(aux,prog,edp)  
 VB(aux,prog,infin)  
 VB(aux,prog,past)  
 VB(aux,prog,past,neg)  
 VB(aux,prog,pres)  
 VB(aux,prog,pres,encl)  
 VB(aux,prog,pres,neg)  
 VB(aux,prog,subjun)  
 VB(aux,prog,subjun,neg)

#### Examples

*He was writing a new book.* VB(aux,prog,past)  
*The theatre is being built right now.* VB(aux,prog,pres)

#### Notes

- The auxiliary may form part of a reduced construction where the lexical verb is absent. In most cases the context indicates that the reduced verb phrase contains an operator realised by an auxiliary of the progressive aspect, as in:

*He was laughing. He really was.* VB(aux,prog,past)

- 's can be the contracted form of *has* and *is*. For example,

*He's started.* VB(aux,perf,pres,encl)  
*He's reading.* VB(aux,prog,pres,encl)

- Similarly *ain't* can be both a perfective and progressive auxiliary:

*He ain't started.* VB(aux,perf,pres,neg)  
*He ain't reading.* VB(aux,prog,pres,neg)

- Be going to* can be both a multi-token semi-auxiliary and a combination of a progressive auxiliary, a lexical verb and a preposition. Compare:

*He was going to Montreal.* VB(aux,prog,past)  
*He was going to stop.*  
 VB(aux,semi,past):1/3 VB(aux,semi,past):2/3 VB(aux,semi,past):3/3

- Be* can also be a lexical verb:

*The meeting is at six.* VB(lex,cop,pres)  
*She's an excellent cook.* VB(lex,cop,pres,encl)  
*It is usually a groom who makes a nuisance of himself.* VB(lex,intr,pres)

- *keep* is tagged as a semi-auxiliary verb followed by an *-ing* participle in sentences such as *She kept crying*. (see section 2.13.1.7). Compare:

*She was crying.* VB(aux,prog,past)  
*She kept crying.* VB(aux,semip,past)

### 2.13.1.4 Perfective auxiliary

The subclass of perfective auxiliary is a closed class. Its only member is *have*. The following tags apply:

VB(aux,perf,edp)  
 VB(aux,perf,infin)  
 VB(aux,perf,infin,encl)  
 VB(aux,perf,ingp)  
 VB(aux,perf,past)  
 VB(aux,perf,past,encl)  
 VB(aux,perf,past,neg)  
 VB(aux,perf,pres)  
 VB(aux,perf,pres,encl)  
 VB(aux,perf,pres,neg)  
 VB(aux,perf,subjun)

#### Examples

*He has written a new novel.* VB(aux,perf,pres)  
*He had already repaired the car.* VB(aux,perf,past)  
*They have gone.* VB(aux,perf,pres)

#### Notes

- 's can be a contracted form of *has* and *is*.

*He's started.* VB(aux,perf,pres,encl)  
*He's swimming.* VB(aux,prog,pres,encl)

- Similarly *ain't* can be a perfective or a progressive auxiliary verb, or a lexical verb:

*He ain't started.* VB(aux,perf,pres,neg)  
*He ain't swimming.* VB(aux,prog,pres,neg)  
*He ain't there.* VB(lex,cop,pres,neg)

- 'd can be a contracted form of *had* and *would*. For example,

*Every day he'd leave the house at 5 a.m.* VB(aux,modal,past,encl)  
*He'd seen her before.* VB(aux,perf,past,encl)

- In reduced sentences the perfective auxiliary *have* is not accompanied by a lexical verb.

*He really has.* VB(aux,perf,pres)  
*Has he really?* VB(aux,perf,pres)

In such constructions the context can be used to determine the right wordclass. Compare:

*He has two children. He really has./Has he really?* VB(lex,intr,pres)  
*He has gone. He really has./Has he really?* VB(aux,perf,pres)

- *have* can also be a lexical verb, as in:

<i>He <u>has</u> two children.</i>	VB(lex,montr,pres)
<i>He <u>hasn't</u> any children.</i>	VB(lex,montr,pres,neg)
<i>Do you <u>have</u> any matches on you?</i>	VB(lex,montr,infin)
<i>He <u>had</u> his office cleaned.</i>	VB(lex,cxtr,past)

*Have* as a lexical verb can be negated or made interrogative by means of the auxiliary *do*

<i>He <u>doesn't have</u> a car.</i>	VB(lex,montr,infin)
<i>Does he <u>have</u> a car?</i>	VB(lex,montr,infin)

### 2.13.1.5 Modal auxiliary

The class of modal auxiliaries is a closed class. Its members are:

1) the central modals: *can, may, must, shall, will*

2) the marginal modals:

<i>dare</i>	<i>need</i>
<i>gotta</i>	<i>ought</i>

multi-token units:

<i>be to</i>	<i>had rather</i>	<i>need to</i>
<i>dare to</i>	<i>have to</i>	<i>ought to</i>
<i>got to</i>	<i>have got to</i>	<i>use(d) to</i>
<i>had better/best</i>	<i>have gotta</i>	

\* elliptical and contracted form of *have got to*

\*\* elliptical form of *have got to*

\*\*\* contracted form of *have got to*

The following tags apply:

VB(aux,modal,edp)  
 VB(aux,modal,infin)  
 VB(aux,modal,ingp)  
 VB(aux,modal,past)  
 VB(aux,modal,past,encl)  
 VB(aux,modal,past,neg)  
 VB(aux,modal,pres)  
 VB(aux,modal,pres,encl)  
 VB(aux,modal,pres,neg)  
 VB(aux,modal,subjun)  
 VB(aux,modal,subjun,neg)

### Examples

<i>You <u>must</u> tell me why she did that.</i>	VB(aux,modal,pres)
<i>He <u>used to</u> like things like these.</i>	VB(aux,modal,past):1/2 VB(aux,modal,past):2/2
<i>He <u>needs to</u> be careful.</i>	VB(aux,modal,pres):1/2 VB(aux,modal,pres):2/2
<i>He <u>dares to</u> ask me that!</i>	VB(aux,modal,pres):1/2 VB(aux,modal,pres):2/2
<i>Does he <u>need to</u> be careful?</i>	VB(aux,modal,infin):1/2 VB(aux,modal,infin):2/2



## Notes

- *Dare* and *need* can also be used as lexical verbs, e.g.

*Are you daring me?* VB(lex,montr,ingp)  
*I need you* VB(lex,montr,pres)

- The modal auxiliaries *dare to*, *have to*, *need to* and *use to* can also be preceded by the auxiliary *do*, e.g.:

*He doesn't dare to come here.* VB(aux,modal,infinitive):1/2 VB(aux,modal,infinitive):2/2  
*He doesn't need to come here.* VB(aux,modal,infinitive):1/2 VB(aux,modal,infinitive):2/2  
*Did he use to drive a car?* VB(aux,modal,infinitive):1/2 VB(aux,modal,infinitive):2/2  
*Does it have to be white?* VB(aux,modal,infinitive):1/2 VB(aux,modal,infinitive):2/2

### 2.13.1.6 Semi-auxiliary

The subclass of semi-auxiliaries is a closed class. Its members are:

<i>appear to</i>	<i>come to</i>	<i>look to</i>
<i>begin to</i>	<i>continue to</i>	<i>seem to</i>
<i>cease to</i>	<i>fail to</i>	<i>start to</i>
<i>be going to</i>	<i>get to</i>	<i>tend to</i>
<i>be gonna</i>	<i>happen to</i>	<i>turn out to</i>

The following tags apply:

VB(aux,semi,edp)  
 VB(aux,semi,imp)  
 VB(aux,semi,infinitive)  
 VB(aux,semi,ingp)  
 VB(aux,semi,past)  
 VB(aux,semi,past,neg)  
 VB(aux,semi,pres)  
 VB(aux,semi,pres,elliptic)  
 VB(aux,semi,pres,enclitic)  
 VB(aux,semi,pres,neg)  
 VB(aux,semi,subjunctive)

### Class characteristics

The class of semi-auxiliaries can be characterised syntactically as follows:

1. Semi-auxiliaries have lexical verbs as their complement. They are complemented by lexical verbs which have the non-finite form. They permit *there*-constructions (cf. CGEL, 3.47):

*Several teams are going to be beaten by England →*  
*There are several teams going to be beaten by England*

2. Semi-auxiliaries resemble authentic auxiliaries in that they constitute one verb phrase with the following lexical verb and therefore permit synonymous passives, such as

*England is going to win the European Cup Final →*  
*The European Cup Final is going to be won by England*

**Examples**

<i>He <u>appeared to</u> be leaving.</i>	VB(aux,semi,past):1/2 VB(aux,semi,past):2/2
<i>They <u>begin to</u> shout.</i>	VB(aux,semi,pres):1/2 VB(aux,semi,pres):2/2
<i>They <u>came to</u> like each other.</i>	VB(aux,semi,past):1/2 VB(aux,semi,past):2/2
<i>The apparatus <u>fails to</u> work silently.</i>	VB(aux,semi,pres):1/2 VB(aux,semi,pres):2/2
<i>They <u>got to</u> know each other.</i>	VB(aux,semi,past):1/2 VB(aux,semi,past):2/2
<i>It <u>happens to</u> take place on a Sunday.</i>	VB(aux,semi,pres):1/2 VB(aux,semi,pres):2/2
<i>It didn't <u>seem to</u> rain.</i>	VB(aux,semi,infin):1/2 VB(aux,semi,infin):2/2
<i>The girl <u>started to</u> run.</i>	VB(aux,semi,past):1/2 VB(aux,semi,past):2/2
<i>They <u>tend to</u> avoid meetings like these.</i>	VB(aux,semi,pres):1/2 VB(aux,semi,pres):2/2
<i>The kidnapping <u>turned out to</u> be bad joke.</i>	VB(aux,semi,past):1/2 VB(aux,semi,past):2/2
<i>The firm <u>ceased to</u> reduce its number of staff.</i>	VB(aux,semi,past):1/2 VB(aux,semi,past):2/2
<i>She <u>was going to</u> take a holiday.</i>	VB(aux,semi,past):1/3 VB(aux,semi,past):2/3 VB(aux,semi,past):3/3

**Notes**

- When it is obvious from the context that one or more tokens of a multi-token semi-auxiliary verb have been ellipted, while at least one other token of the unit is present, the feature 'ellipt' is assigned. For example,

<i><u>Going to</u> swim?</i>	VB(aux,semi,past,ellipt):1/2 VB(aux,semi,past,ellipt):2/2
<i><u>Going to</u> swim, was he?</i>	VB(aux,semi,past,ellipt):1/2 VB(aux,semi,past,ellipt):2/2

- In constructions as *be bound to*, *be certain to*, *be meant to*, *be due to*, *be liable to*, *be sure to*, *be (un)likely to*, the tokens *bound*, *certain*, *meant*, *due*, *liable*, *sure*, *(un)likely* are tagged as general adjectives.

- seem* can also be a copula. For example,

<i>They <u>seemed</u> very friendly.</i>	VB(lex,cop,past)
--	------------------

- appear* can also be a copula or an intransitive lexical verb:

<i>He <u>appeared</u> drunk</i>	VB(lex,cop,past)
<i>He <u>appeared</u> in Nijmegen</i>	VB(lex,intr,past)

- When a verb is followed by a direct object instead of a particle *to* plus verbal complement realised by an infinitive verb, it should be tagged as a monotransitive lexical verb. Compare:

<i>They <u>start to</u> eat.</i>	VB(aux,semi,pres):1/2 VB(aux,semi,pres):2/2
<i>They <u>continued to</u> dance.</i>	VB(aux,semi,past):1/2 VB(aux,semi,past):2/2
<i>And then they <u>start</u> the game.</i>	VB(lex,montr,pres)
<i>The couple <u>continued</u> their dancing.</i>	VB(lex,montr,past)

**2.13.1.7 Semi-auxiliary followed by -ing participle**

The class of semi-auxiliaries that are followed by *-ing* participles is a closed class. Its members are:

<i>begin</i>	<i>finish</i>	<i>start</i>
<i>continue</i>	<i>keep</i>	<i>stop</i>

Multi-token units:

*carry on*  
*go on*

*keep on*  
*set about*

*start out*

The following tags apply:

VB(aux,semip,edp)  
VB(aux,semip,imp)  
VB(aux,semip,infin)  
VB(aux,semip,ingp)  
VB(aux,semip,past)  
VB(aux,semip,pres)  
VB(aux,semip,subjun)

### Class characteristics

Syntactically, semi-auxiliaries that are followed by *-ing* participles can be characterised as follows:

1. the verb resembles a semi-auxiliary in that it has a lexical verb as its complement.
2. the auxiliary is complemented by a lexical verb which has the *-ing* participle form and constitutes one verb phrase with it.

Semantically, the auxiliary should be aspectual in nature: it should refer to the beginning, continuation or end of an action.

### Examples

<i>The chair <u>began</u> interrupting the speaker.</i>	VB(aux,semip,past)
<i>The couple <u>continued</u> dancing until midnight.</i>	VB(aux,semip,past)
<i>The children will <u>finish</u> playing when the sun sets.</i>	VB(aux,semip,infin)
<i><u>Keep</u> smiling.</i>	VB(aux,semip,imp)
<i>Don't you <u>start</u> swearing too.</i>	VB(aux,semip,infin)
<i><u>Stop</u> shouting; I'm not deaf.</i>	VB(aux,semip,imp)
<i><u>Carry on</u> working, please.</i>	VB(aux,semip,imp):1/2 VB(aux,semip,imp):2/2
<i><u>They went on</u> pestering the poor child.</i>	VB(aux,semip,past):1/2 VB(aux,semip,past):2/2
<i><u>Keep on</u> walking, gentlemen.</i>	VB(aux,semip,imp):1/2 VB(aux,semip,imp):2/2
<i>The officer <u>starts out</u> walking his dog.</i>	VB(aux,semip,pres):1/2 VB(aux,semip,pres):2/2
<i>With the tent perfectly in order, they <u>set about</u> finding water.</i>	VB(aux,semip,past):1/2 VB(aux,semip,past):2/2

### Notes

- If the auxiliary is followed by an infinitive instead of by an *-ing* participle, it is tagged as semi-auxiliary. Compare:

<i>They <u>began to</u> dance across the stage.</i>	VB(aux,semi,past):1/2 VB(aux,semi,past):2/2
<i>They <u>began</u> dancing across the stage.</i>	VB(aux,semip,past)

If it is followed by a direct object, it is tagged as a (monotransitive) lexical verb. For example:

<i>And then they <u>start</u> the game.</i>	VB(lex,montr,pres)
<i>The couple <u>continued</u> their dancing.</i>	VB(lex,montr,past)

### 2.13.2 Lexical verb

The tagging of the subclass of lexical verb in a given utterance is syntactically motivated: i.e. it is based on the actual presence of any objects and/or complements in that utterance. The tagging of a lexical verb is as follows:

1. In declarative and interrogative sentences where the verb phrase is in the active voice,

— a verb is tagged as 'intransitive' if there are no objects and/or complements; e.g.

*Hackett smiled.* VB(lex,intr,past)

— a verb is tagged as 'copular' if it is accompanied by a subject complement; e.g.

*She didn't sound worried.* VB(lex,cop,infin)  
*Yes, silly it is.* VB(lex,cop,pres)

— a verb is tagged as 'monotransitive' if a direct object is present; e.g.

*Hackett faked weariness.* VB(lex,montr,past)  
*That I cannot hear.* VB(lex,montr,infin)

— a verb is tagged as 'dimono-transitive' if there is an indirect object but no direct object; e.g.

*Any doctor will tell you.* VB(lex,dimontr,infin)  
*Their secretary, he told me, had been fired.* VB(lex,dimontr,past)

— a verb is tagged as 'ditransitive' if it is accompanied by both an indirect and a direct object; e.g.

*I did you no harm.* VB(lex,ditr,past)

— a verb is tagged as 'complex transitive' if there are a direct object and an object complement present; e.g.

*I wrapped the canvas around her.* VB(lex,cxtr,past)  
*Hackett had discarded as purposeful fictions most of the titbits James had let drop about himself.* VB(lex,cxtr,edp)

2. In regular declarative and interrogative sentences where the verb phrase is in the passive voice, the tagging of the lexical verb is the same as it would have been if the sentence had been in the active voice, e.g.

*He was appointed president.* VB(lex,cxtr,edp)  
*She has been awarded a scholarship.* VB(lex,ditr,edp)

Verbs that are normally intransitive in the active voice are tagged as monotransitive when they occur in passive verb phrases, e.g.

*Somebody slept in my bed.* VB(lex,intr,past)  
*My bed has been slept in.* VB(lex,montr,edp)

3. 'Prepositional verbs' as well as 'phrasal-prepositional verbs', i.e. verbs accompanied by a phrasal preposition, are tagged as intransitive in active sentences, but as monotransitive in passive sentences, in accordance with 2. above, e.g.

*Everybody looked at the girl.* VB(lex,intr,past)  
*The girl was looked at by everybody.* VB(lex,montr,edp)

4. 'Phrasal verbs', i.e. verbs accompanied by a phrasal adverb, are normally tagged as monotransitive in active sentences when they take a direct object, e.g.

*His father drew up a new will.* VB(lex,montr,past)  
*A new will was drawn up months ago.* VB(lex,montr,edp)  
*Did you make up this story?* VB(lex,montr,infin)

Again, such verbs are monotransitive in passive sentences.

5. In cleft sentences, the verb *be* is always tagged as intransitive. For example:

*It was in Belfast that he met his friend.* VB(lex,intr,past)

6. In zero relative clauses, the zero object is relevant to the assignment of the main verb tag. For example, given the sentence

*This is the patient (who was) scanned yesterday*

the verb *scanned* is tagged as VB(lex,montr,edp) irrespective of whether or not the pronoun is actually present.

### Notes

- Some verbs can be monotransitive lexical verbs or semi-auxiliaries or semi-auxiliaries followed by an *-ing* participle. For example,

*They started to dance* VB(aux,semi,past):1/2 VB(aux,semi,past):2/2  
*They kept dancing.* VB(aux,semip,past)

But:

*They started their dance.* VB(lex,montr,past)  
*I've kept your seat.* VB(lex,montr,edp)

- In sequences of lexical verb plus noun phrase plus non-finite construction, the lexical verb may be either ditransitive or monotransitive, according to whether the NP functions as indirect object to the lexical verb, or as subject of the verb in a non-finite, infinitive clause.

*They asked John to do the job.* VB(lex,ditr,past)  
 cf. *What did they ask John?*

*She asked me to wash the dishes.* VB(lex,ditr,past)  
 cf. *What did she ask me?*

*He expects her to call me.* VB(lex,montr,pres)  
 cf. *What does he expect?*

In some cases the *wh*-question does not work, e.g.:

*He persuaded Mary to see a doctor.*

cf. \*What did he persuade?  
\*What did he persuade Mary?

In such cases the NP (*Mary*) is considered to be the explicit indirect object of the verb (*persuade*), while the subject of the infinitive (*to see*) is implicit. As a consequence, such verbs are tagged as ditransitive:

*He persuaded Mary to see a doctor.* VB(lex,ditr,past)

### 2.13.2.1 Intransitive verb

The subclass of intransitive lexical verbs is an open class (cf. CGEL 16.19). It contains both single tokens and multi-token units. Multi-token lexical verbs are lexically frozen. They have at least one token which is morpho-syntactically inflexible, and which cannot appropriately be classified as an individual item. Examples are *follow suit* and *let go*.

The following tags apply:

VB(lex,intr,edp)  
VB(lex,intr,imp)  
VB(lex,intr,infin)  
VB(lex,intr,ingp)  
VB(lex,intr,past)  
VB(lex,intr,past,neg)  
VB(lex,intr,pres)  
VB(lex,intr,pres,encl)  
VB(lex,intr,pres,neg)  
VB(lex,intr,subjun)

#### Class characteristics

A verb is tagged as intransitive ('intr') if it does not have any (direct or indirect) object or (subject or object) complement.

#### Examples

*He smiled.* VB(lex,intr,past)  
*He walked slowly towards the door.* VB(lex,intr,past)  
*The younger men were moaning.* VB(lex,intr,ingp)

#### Notes

- Prepositions should be tagged as phrasal prepositions if they form part of 'prepositional verbs' or 'phrasal-prepositional verbs', that is, if there is a strong collocational link between the verb and the preposition, while the lexical verb and preposition together form a semantic unit. Verbs accompanied by a phrasal preposition are always tagged as intransitive verbs in active sentences. In passive sentences, however, they receive the monotransitive tag.

*He used to rely on me completely.* VB(lex,intr,infin) PREP(phras)  
*Did you really apply for that job?* VB(lex,intr,infin) PREP(phras)  
*Stop laughing at him immediately!* VB(lex,intr,ingp) PREP(phras)  
*He was laughed at continuously.* VB(lex,montr,edp) PREP(phras)  
*He cannot be relied on.* VB(lex,montr,edp) PREP(phras)  
*The position has not been applied for.* VB(lex,montr,edp) PREP(phras)

### 2.13.2.2 Copular verb

The class of copular verbs is an open class (cf. CGEL 10.8). Examples of copular verbs are:

<i>appear</i>	<i>go</i>	<i>prove</i>	<i>stay</i>
<i>be</i>	<i>grow</i>	<i>remain</i>	<i>taste</i>
<i>become</i>	<i>keep</i>	<i>seem</i>	<i>turn</i>
<i>feel</i>	<i>look</i>	<i>smell</i>	<i>wax</i>
<i>get</i>	<i>make</i>	<i>sound</i>	

Multi-token copular verbs are those combinations of tokens which are lexically frozen and which contain at least one token which is morpho-syntactically inflexible, while it is difficult to classify this token as a member of a separate wordclass. An example is:

*He turned out extremely dangerous.*      VB(lex,cop,past):1/2 VB(lex,cop,past):2/2

The following tags apply:

VB(lex,cop,edp)  
 VB(lex,cop,imp)  
 VB(lex,cop,infin)  
 VB(lex,cop,ingp)  
 VB(lex,cop,past)  
 VB(lex,cop,past,neg)  
 VB(lex,cop,pres)  
 VB(lex,cop,pres,encl)  
 VB(lex,cop,pres,neg)  
 VB(lex,cop,subjun)  
 VB(lex,cop,subjun,neg)

#### Class characteristics

A copular lexical verb is a verb which is complemented by a subject complement.

#### Examples

<i>They <u>were</u> in a hurry.</i>	VB(lex,cop,past)
<i>He <u>seemed</u> under the impression that it was alright.</i>	VB(lex,cop,past)
<i>The reason <u>is</u> that he is a crook.</i>	VB(lex,cop,pres)
<i>My idea of a good evening <u>was</u> going to the casino.</i>	VB(lex,cop,past)
<i>Mary was <u>getting</u> tired of it.</i>	VB(lex,cop,ingp)
<i>Her husband had <u>gone</u> quite tense and jumpy.</i>	VB(lex,cop,edp)
<i>They had <u>dropped</u> asleep while waiting.</i>	VB(lex,cop,edp)
<i>He <u>fell</u> asleep.</i>	VB(lex,cop,past)

#### Note

- If a verb can also be used intransitively with a similar meaning, it is tagged as an intransitive verb, not as a copula: e.g. *act as*, *count as*, *drop down*, *end up*, *function as*, *make up*, *pose as*, *serve as*.

### 2.13.2.3 Monotransitive verb

The class of monotransitive verbs is an open class (cf. CGEL 16.26). The class contains both single tokens and multi-token units. Multi-token monotransitive verbs are lexically fixed. They contain at least one token which is morpho-syntactically fixed too, and which cannot appropriately be accommodated in a separate wordclass. Examples of multi-token units are: *get rid of*, *get wind of*, *give rise to*, *keep pace with*, *make sure*, *mass mobilise*, *short circuit*, *take hold of*.

The following tags apply:

VB(lex,montr,edp)  
 VB(lex,montr,imp)  
 VB(lex,montr,infin)  
 VB(lex,montr,ingp)  
 VB(lex,montr,past)  
 VB(lex,montr,past,encl)  
 VB(lex,montr,past,neg)  
 VB(lex,montr,pres)  
 VB(lex,montr,pres,encl)  
 VB(lex,montr,pres,neg)  
 VB(lex,montr,subjun)

### Class characteristics

A verb is tagged as monotransitive if it has only a direct object, and not an indirect object or another complement.

### Examples

<i>He <u>possesses</u> a magnificent house in France.</i>	VB(lex,montr,pres)
<i>He <u>had</u> the best qualities of all.</i>	VB(lex,montr,past)
<i>She <u>knows</u> where they are.</i>	VB(lex,montr,pres)
<i>I <u>wonder</u> why Jane's done that.</i>	VB(lex,montr,pres)
<i>Frank <u>loathes</u> swimming in the Atlantic Ocean.</i>	VB(lex,montr,pres)

### Notes

- No multi-token unit is: *find (to be) wanting*
- In elliptical constructions, no feature is used to express the ellipsis:
 

<i>They bought a lot more than anyone of us could <u>buy</u>.</i>	VB(lex,montr,infin)
---	---------------------
- In coordinations lexical verbs should be tagged as monotransitive if they share a direct object with another monotransitive verb. An example is:
 

<i>He <u>saw</u>, but did not buy the watch.</i>	VB(lex,montr,past)
--	--------------------
- Generally, a direct object in an active (declarative or interrogative) construction can function as the subject in a corresponding passive equivalent. Exceptions, which can be monotransitive verbs in active sentences, but which cannot be used in corresponding passive forms are *become, equal, fit, have, hold, lack, match, possess, suit*. (Cf. CGEL 16.27).
- Noun phrases following verbs may be adverbials instead of direct objects. In such contexts, the verb should be tagged as intransitive. Examples are:
 

<i>She <u>called</u> this morning.</i>	VB(lex,intr,past)
<i>He <u>spoke</u> that way.</i>	VB(lex,intr,past)
- prepositions which form part of prepositional verbs or phrasal-prepositional verbs are tagged as phrasal prepositions: PREP(phras), that is: if there is a strong collocational link between the verb and the preposition, and if they together form a semantic unit. The



same goes for adverbs in phrasal or phrasal-prepositional verbs, which is tagged as phrasal adverbs. Examples are:

*He compared himself with his father.* VB(lex,montr,past) PREP(phras)  
*The policeman wrote the number down* VB(lex,montr,past) ADV(phras)

### 2.13.2.4 Ditransitive verb

The class of ditransitive verbs is an open class. The following tags apply:

VB(lex,ditr,edp)  
 VB(lex,ditr,imp)  
 VB(lex,ditr,infin)  
 VB(lex,ditr,ingp)  
 VB(lex,ditr,past)  
 VB(lex,ditr,pres)  
 VB(lex,ditr,subjun)

#### Class characteristics

A verb is tagged as ditransitive ('ditr') if it has both a direct object and an indirect object.

#### Examples

*He gave John a book.* VB(lex,ditr,past)  
*Any worry Cynthia may have caused you will now have disappeared.* VB(lex,ditr,edp)  
*He convinced me that the alternative was more realistic.* VB(lex,ditr,past)  
*She brought the librarian a most wonderful edition of Hamlet.* VB(lex,ditr,past)  
*Tell me another story.* VB(lex,ditr,imp)

#### Notes

- Compare:

*He built himself a house.* VB(lex,ditr,past)  
*He built a house for himself.* VB(lex,montr,past) PREP(ge)

- A non-finite verb which is part of a non-finite clause is tagged in the same way as when it forms part of a finite clause:

*The book given her by Peter was an excellent choice.* VB(lex,ditr,edp)

- In cases where the NP following the verb is considered to be the explicit indirect object of the verb (*persuade, tell, force, etc.*), while the subject of the infinitive direct object is implicit, the verb is tagged as ditransitive. If the direct object is absent, the verb should be tagged as dimono-transitive (see section 2.13.2.5), e.g.

*He persuaded Mary to see a doctor.* VB(lex,ditr,past)  
*He forced me to sign the letter.* VB(lex,ditr,past)  
*I told John the story* VB(lex,ditr,past)  
*He persuaded Mary.* VB(lex,dimonstr,past)  
*He forced me.* VB(lex,dimonstr,past)  
*I told John* VB(lex,dimonstr,past)  
*Their secretary, he told me, had been fired.* VB(lex,dimonstr,past)

### 2.13.2.5 Dimono-transitive verb

The class of dimono-transitive verbs is an open class. The following tags apply:

VB(lex,dimontr,edp)  
 VB(lex,dimontr,imp)  
 VB(lex,dimontr,infin)  
 VB(lex,dimontr,ingp)  
 VB(lex,dimontr,past)  
 VB(lex,dimontr,pres)  
 VB(lex,dimontr,subjun)

### Class characteristics

When a ditransitive verb, which requires both a direct and an indirect object, is found in a context where the direct object is absent, and only the indirect object is present, it is tagged as 'dimono-transitive' (dimontr).

### Examples

<i>The deal will come through, I <u>grant</u> you.</i>	VB(lex,dimontr,pres)
<i>Had I been <u>allowed</u>, I would have joined the airforce.</i>	VB(lex,dimontr,edp)
<i>Mamoulian had <u>told</u> him about Whitehead and the dogs.</i>	VB(lex,dimontr,edp)
<i>Have you <u>asked</u> her?</i>	VB(lex,dimontr,edp)

### Notes

- Normally, a ditransitive verb should be tagged as dimono transitive if it occurs with an indirect object, but no direct object. Note that the indirect object may be followed by an adverbial realized by a noun phrase:

*Frances told me this morning.* VB(lex,dimontr,past)

Compare:

<i>He <u>wrote</u> me</i>	VB(lex,dimontr,past).
<i>He <u>wrote</u> to me.</i>	VB(lex,intr,past) PREP(ge)

## 2.13.2.6 Complex transitive verb

The class of complex transitive verbs is an open class.

### Class characteristics

A complex transitive verb has a direct object and an object complement:

1. The object complement is semantically related to the direct object of the lexical verb in such a way that it attributes a state or quality to the referent of the direct object. This semantic relation can generally be paraphrased in a finite sentence with a copular verb, where the subject corresponds with the direct object and the subject complement with the (original) object complement. *He kept the bookshelf in place → The bookshelf is in place.*
2. The object complement is obligatory if it is realised by a locative PP. Two examples are: *He put the car into the garage and He saw her to the car.* If the object complement is realised by an adjective phrase, a noun phrase, or an as-phrase, it is sometimes omissible. Examples are: *He painted the door green. They appointed him chairman, and He introduced the girl as his wife.* Note that in *he painted the door, they appointed him and he introduced the girl*, the verbs are monotransitive.
3. The *where*-criterion: if the locative PP can be queried by means of *where*, the verb is not complex-transitive and the phrase is not an object complement, but an adverbial. An example of a complex-transitive verb is: *He buried his nose in his books → \*Where did he bury his nose? Cf. They buried the body in the churchyard → Where did they bury the body?*

### Semantic subclasses of complex transitives

- movement/location (literal)

Examples of such complex-transitive verbs are: *put, pull, get, let, bring, take, push, drag, lead, tuck, stow, slip, jerk, drag, bang, shove, pitch, shoot, lay, keep*

Further examples are:

<i>He <u>buried</u> his nose in his books.</i>	VB(lex,cxtr,past)
<i>She <u>held</u> the lens in place.</i>	VB(lex,cxtr,past)
<i>After the party, they <u>saw</u> their guests out.</i>	VB(lex,cxtr,past)
<i><u>See</u> Mr. Lawson to the car, will you?</i>	VB(lex,cxtr,imp)

- movement/state (metaphorical)

Examples of such complex transitive verbs are: *render, drive, turn, form, scare, terrify*

Further examples are:

<i>They <u>rendered</u> it useless</i>	VB(lex,cxtr,past)
<i>She <u>drove</u> him barking mad.</i>	VB(lex,cxtr,past)
<i>The shriek <u>terrified</u> them stiff.</i>	VB(lex,cxtr,past)

- perception

Examples of such complex transitive verbs are: *consider, conclude, judge, imagine, call*

Further examples are:

<i>I <u>consider</u> his decision very foolish.</i>	VB(lex,cxtr,pres)
<i>They <u>judged</u> it insufficient.</i>	VB(lex,cxtr,past)
<i>"Don't you <u>call</u> me a liar!" he roared.</i>	VB(lex,cxtr,infin)

- volition

Examples of such complex transitive verbs are: *have, want*

An examples is:

<i>I <u>want</u> him in my office, now!</i>	VB(lex,cxtr,pres)
---	-------------------

### Notes

- The object complement may be realized by noun phrase, adjective phrase, adverb phrase or a prepositional phrase. A special instance of the latter case is found when the preposition is *as*; this is always tagged as a phrasal preposition. Examples of such complex transitive verbs are: *introduce, employ, strike, class, see, regard, describe, suggest, view, imagine, exhibit, conjecture, take*. Examples are:

<i>The student <u>introduced</u> his friend <u>as</u> Peter Smith.</i>	VB(lex,cxtr,past) PREP(ge)
<i>They <u>regarded</u> the discussion <u>as</u> closed.</i>	VB(lex,cxtr,past) PREP(phras)
<i>She could mistake him <u>for</u> her father.</i>	VB(lex,cxtr,infin) PREP(phras)

- A non-finite verb which introduces a non-finite clause is tagged in the same way as when it forms part of a finite clause. An example is:

*I met a girl called Frances*                      VB(lex,cxtr,edp)

## 2.14 Miscellaneous

The 'miscellaneous' class serves to accommodate a variety of items that do not belong to any of the other major wordclasses. Five subclasses are distinguished: discourse item, foreign word or expression, interjection, suffix and prefix. The following tags apply:

MISC(discourse)  
MISC(foreign)  
MISC(interjec)  
MISC(prefix)  
MISC(suffix)

### 2.14.1 Discourse item

The subclass of discourse items is an open class. Examples of discourse items are:

#### a. formulae (see CGEL section 11.54)

<i>abracadabra</i>	<i>gangway</i>	<i>please</i>
<i>adieu</i>	<i>goodbye</i>	<i>pray</i>
<i>ahoy</i>	<i>goodday</i>	<i>prithiee</i>
<i>alack</i>	<i>goodnight</i>	<i>regards</i>
<i>alas</i>	<i>hallo</i>	<i>shalom</i>
<i>alleluia</i>	<i>hello</i>	<i>sorry</i>
<i>avaunt</i>	<i>here</i>	<i>ta</i>
<i>bravo</i>	<i>hi</i>	<i>ta-ta</i>
<i>bye</i>	<i>hooray</i>	<i>tally-ho</i>
<i>bye-bye</i>	<i>halleluja</i>	<i>thanks</i>
<i>cheerio</i>	<i>hullo</i>	<i>there</i>
<i>cheers</i>	<i>huzza</i>	<i>timber</i>
<i>congratulations</i>	<i>like</i>	<i>wotcher</i>
<i>encore</i>	<i>okay</i>	<i>welcome</i>
<i>eureka</i>	<i>OK</i>	
<i>farewell</i>	<i>pardon</i>	

multi-token units:

<i>I prithiee</i>	<i>here you are</i>	<i>right on</i>
<i>all the best</i>	<i>hey presto</i>	<i>see you</i>
<i>bye bye</i>	<i>hip hip hooray</i>	<i>so long</i>
<i>hands off</i>	<i>mind you</i>	<i>steady on</i>
<i>hands up</i>	<i>never mind</i>	<i>there you are</i>
<i>hear, hear</i>	<i>open sesame</i>	<i>wakey wakey</i>

**b. Responsive phrases:**

<i>aye</i>	<i>no</i>	<i>yea</i>
<i>nay</i>	<i>nope</i>	<i>yeah</i>
<i>nix</i>	<i>roger</i>	<i>yes</i>

**c. Expletives:**

<i>bags</i>	<i>crap</i>	<i>goody</i>
<i>begad</i>	<i>crikey</i>	<i>gosh</i>
<i>ballocks</i>	<i>cripes</i>	<i>heavens</i>
<i>blimey</i>	<i>damn</i>	<i>heck</i>
<i>bollocks</i>	<i>damnation</i>	<i>hell</i>
<i>balls</i>	<i>darn</i>	<i>humbug</i>
<i>bingo</i>	<i>fiddlesticks</i>	<i>hosanna</i>
<i>blast</i>	<i>fuck</i>	<i>knickers</i>
<i>bother</i>	<i>gad</i>	<i>nuts</i>
<i>botheration</i>	<i>golly</i>	<i>rats</i>
<i>bull</i>	<i>goodness</i>	<i>shit</i>
<i>bullshit</i>	<i>goodie</i>	<i>shucks</i>

**multi-token units:**

<i>bloody hell</i>	<i>by Jove</i>	<i>lo and behold</i>
<i>by golly</i>	<i>fucking hell</i>	

**Class characteristics**

The class of discourse items comprises greetings and other formulae, expletives and responsive phrases. They have an interactive function in discourse and have no referring content.

**Note**

- Although the discourse items constitute an open class, their membership is kept as limited as possible, especially with respect to multi-token units. Whenever a feasible tagging is possible, word sequences with a discursual function are tagged as separate items; for example,

<i>big deal</i>	<i>good evening</i>	<i>many thanks</i>
<i>bugger off</i>	<i>good health</i>	<i>Merry Christmas</i>
<i>Christ</i>	<i>good morning</i>	<i>my God</i>
<i>damn it</i>	<i>good night</i>	<i>my mistake</i>
<i>damn you</i>	<i>Happy Birthday</i>	<i>no problem</i>
<i>dear me</i>	<i>Happy New Year</i>	<i>pray God</i>
<i>don't mention it</i>	<i>hard lines</i>	<i>thank God</i>
<i>excuse me</i>	<i>hard luck</i>	<i>very well</i>
<i>for Christ's sake</i>	<i>have a nice day</i>	<i>watch it</i>
<i>for God's sake</i>	<i>how are you?</i>	<i>watch out</i>
<i>fuck it</i>	<i>how do you do?</i>	<i>well done</i>
<i>fuck you</i>	<i>I'm sorry</i>	<i>you're welcome</i>
<i>get lost</i>	<i>I beg your pardon</i>	<i>you know</i>
<i>glad to meet you</i>	<i>I mean</i>	<i>you see</i>
<i>go away</i>	<i>I see</i>	
<i>God knows</i>	<i>Jesus Christ</i>	
<i>good afternoon</i>	<i>look out</i>	

### 2.14.2 Foreign word or expression

The subclass of foreign words and expressions is an open class.

#### Class characteristics

Foreign words or expressions are tagged MISC(foreign). The tag also applies to those (parts of) foreign proper names that start with a small letter.

#### Examples

<i>Peter <u>de</u> Vries</i>	MISC(foreign)
<i>He lives in Broek <u>op</u> Lange Dijk.</i>	MISC(foreign)
<i>His '<u>hoofdpijn</u>' wasn't as bad as we expected.</i>	MISC(foreign)

#### Note

- (Parts of) foreign proper names are tagged N(sing) if the tokens start with capitals. Those tokens which form part of a foreign proper name, and which start with a small letter, are tagged as MISC(foreign). For example,

<i>Peter de Vries</i>	N(sing) MISC(foreign) N(sing)
-----------------------	-------------------------------

### 2.14.3 Interjection

The subclass of interjections is an open class. Examples of interjections are:

<i>ah</i>	<i>hey</i>	<i>sh</i>
<i>aha</i>	<i>h'm</i>	<i>shooh</i>
<i>ahem</i>	<i>hm</i>	<i>strewth</i>
<i>atishoo</i>	<i>hum</i>	<i>struth</i>
<i>bah</i>	<i>humph</i>	<i>tut</i>
<i>boo</i>	<i>jiminy</i>	<i>tut-tut</i>
<i>boohoo</i>	<i>mmm</i>	<i>ugh</i>
<i>bosh</i>	<i>oh</i>	<i>uh</i>
<i>bowwow/bow-wow</i>	<i>oof</i>	<i>uh-uh</i>
<i>cor</i>	<i>ooh</i>	<i>uhuh</i>
<i>ee</i>	<i>oops</i>	<i>uhm</i>
<i>eh</i>	<i>ouch</i>	<i>uhum</i>
<i>er</i>	<i>ow</i>	<i>vroom</i>
<i>erm</i>	<i>pah</i>	<i>vroom-vroom</i>
<i>faugh</i>	<i>peekaboo</i>	<i>whew</i>
<i>fie</i>	<i>peepbo</i>	<i>whoa</i>
<i>gee</i>	<i>phew</i>	<i>whoopee</i>
<i>ha</i>	<i>phoey</i>	<i>woof</i>
<i>ha-ha</i>	<i>pooh</i>	<i>wow</i>
<i>haw-haw</i>	<i>pooh-pooh</i>	<i>yippee</i>
<i>heighho</i>	<i>pssst</i>	<i>yuk</i>
<i>hem</i>	<i>rat-ta-ta-ta-ta</i>	

multi-token units:

<i>gee whiz</i>	<i>ha ha</i>	<i>rat ta ta ta ta</i>
<i>bow wow</i>	<i>haw haw</i>	<i>vroom vroom</i>

### Class characteristics

Interjections are primarily used to express emotions. On the whole they have a different morphological and sometimes even phonological structure from other lexical items. Some of them are onomatopaeic in character.

### 2.14.4 Prepositioned bound morpheme (prefix)

The subclass of prepositioned bound morphemes is a closed class. Its members are:

<i>pre</i>	<i>mid</i>	<i>semi</i>
<i>pro</i>	<i>neo</i>	<i>supra</i>
<i>post</i>	<i>sub</i>	<i>sub</i>
<i>macro</i>	<i>hyper</i>	<i>counter</i>
<i>anti</i>	<i>non</i>	<i>super</i>
<i>counter</i>	<i>self</i>	<i>ultra</i>
<i>ante</i>	<i>ex</i>	<i>cross</i>

Measurements:

<i>tera</i>	<i>deca</i>	<i>nano</i>
<i>giga</i>	<i>deci</i>	<i>pico</i>
<i>mega</i>	<i>centi</i>	<i>femto</i>
<i>kilo</i>	<i>milli</i>	<i>atto</i>
<i>hecto</i>	<i>micro</i>	

### Class characteristics

Bound morphemes are generally spelled as part of words or are separated from them by means of hyphens. However, in cases where they are separated from a free morpheme by means of a space or bracket, they are tagged as independently used morphemes.

### Examples

<i>sub contractor</i>	MISC(prefix)
<i>post war</i>	MISC(prefix)
<i>semi professional</i>	MISC(prefix)
<i>neo classical</i>	MISC(prefix)

### Note

- Prefixes are only tagged as miscellaneous prefixes if they are separated from their headword by a blank. They may or may not be accompanied by a hyphen. For example,

<i>pre exam</i>	MISC(prefix)
<i>pre- and postwar subcultures</i>	MISC(prefix)

If the prefix is connected to a free morpheme by means of a hyphen, it forms one token with this free morpheme. Compare:

<i>pre-exam</i>	N(sing)
<i>pre exam</i>	MISC(prefix) N(sing)

### 2.14.5 Postpositioned bound morpheme (suffix)

The subclass of postpositioned bound morphemes is a closed class. Its members are: *wise*, *like*, and (s).

### Class characteristics

Bound morphemes are generally spelled as part of words or are separated from them by means of hyphens. However, in cases where they are separated from a free morpheme by means of a space or bracket, they are tagged as independently used morphemes.

### Examples

<i>house</i> <u>(s)</u>	MISC(suffix)
<i>pairs</i> <u>wise</u>	MISC(suffix)
<i>friendly</i> <u>like</u>	MISC(suffix)

### Note

- Suffixes are only tagged as miscellaneous suffixes if they are separated from their headword by a blank or by an opening bracket. If the suffix is connected to a free morpheme by means of a hyphen, it is not tagged as a separate token. Compare:

<i>friendly-like</i>	ADJ(ge,pos)
<i>friendly like</i>	ADJ(ge,pos) MISC(suffix)

*like* could also be a discourse item, but would then probably be preceded by a comma.

## 2.15 Genitive marker

The class of genitive markers is a closed class. Its members are ' and 's. They receive the major wordclass label GENM. There is no additional feature information. The tag that applies is GENM.

### Examples

<i>He was very inquisitive about other <u>folks'</u> affairs.</i>	GENM
<i>John<u>'s</u> car is a Buick.</i>	GENM

## 2.16 Punctuation

Punctuation receives the major wordclass label PUNC. Additional feature information relates to subclass (obrack, cbrack, colon, comma, dash, ellip, exm, oquo, cquo, per, qm, scolon, other). The tags that apply are the following:

PUNC(obrack)	PUNC(ellip)	PUNC(qm)
PUNC(cbrack)	PUNC(exm)	PUNC(scolon)
PUNC(colon)	PUNC(oquo)	
PUNC(comma)	PUNC(cquo)	PUNC(other)
PUNC(dash)	PUNC(per)	

### Class characteristics

The class of punctuation is a closed, but non-enumerable class. Usually punctuation is contracted with the word preceding or following it.

### Note

- The subclass of 'other' punctuation forms a default category for any punctuation that does not belong to any of the other subclasses. The tags apply as follows:

TAG	explanation	symbol	SGML code
PUNC(obrack)	opening bracket:	( [	&lpar; &lsqb;



		{	&lcurly;
		<	&lt;
PUNC(cbrack)	closing bracket:	)	&rpar;
		]	&rsqb;
		}	&rcub;
		>	&gt;
PUNC(oquo)	opening quote:	‘	&lquo;
		“	&ldquo;
PUNC(cquo)	closing quote:	’	&rquo;
		”	&rdquo;
PUNC(colon)	colon:	:	colon;
PUNC(comma)	comma:	,	&comma;
PUNC(dash)	dash:	—	&dash;
PUNC(exm)	exclamation mark:	!	&excl;
PUNC(per)	period:	.	&period;
PUNC(qm)	question mark:	?	&quest;
PUNC(scolon)	semi-colon:	;	&semi;
PUNC(ellip)	ellipsis:	...	&hellip;
PUNC(other)	other types of punctuation, incl.	• ♦	

### 3. REFERENCES

- Aarts, J. and F. Aarts, 1982. *English Syntactic Structures*. Prentice Hall.
- De Haan, P. and H. van Halteren. 1997. *The TOSCA-ICLE Tagset: Software manual*. Nijmegen: University of Nijmegen
- Oostdijk, N. 1990. *The SEU tagset - manual*. Nijmegen: University of Nijmegen.
- Oostdijk, N., 1991. *Corpus Linguistics and the Automatic Analysis of English*. Amsterdam-Atlanta, GA. Rodopi.
- Quirk, R., S. Greenbaum, G. Leech and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

## APPENDIX 1: ALPHABETICAL LIST OF WORDCLASS AND FEATURE LABELS

### Major wordclasses (16)

ADJ	adjective
ADV	adverb
ART	article
CONJUNC	conjunction
EXTHERE	existential <i>there</i>
GENM	genitive marker
MISC	miscellaneous
N	noun
NADJ	nominal adjective
NUM	numeral
PREP	preposition
PROFM	proform
PRON	pronoun
PRTCL	particle
PUNC	punctuation
VB	verb

### Features

antit	anticipatory <i>it</i>	PRON
ass	assertive	PRON
aux	auxiliary	VB
card	cardinal	NUM
cbrack	closing bracket	PUNC
clause	clause	PROFM
cleft	cleft <i>it</i>	PRON
collect	collective	N
colon	colon	PUNC
comma	comma	PUNC
comp	comparative	ADJ; ADV; NADJ
conj	conjoin	PROFM
connec	connective	ADV
coord	coordinating	CONJUNC
cop	copula	VB
cquo	closing quote	PUNC
cctr	complex transitive	VB
dash	dash	PUNC
def	definitive	ART
dem	demonstrative	PRON
dimontr	dimono-transitive	VB
discourse	discourse	MISC
ditr	ditransitive	VB
do	<i>do</i>	VB
edp	<i>-ed</i> participle	ADJ; NADJ; VB
ellip	ellipsis	PUNC
ellipt	elliptical	VB
encl	enclitic	PRON; VB
exclam	exclamatory	PRON
exm	exclamation mark	PUNC
for	particle <i>for</i>	PRTCL

foreign	foreign	MISC
frac	fractional	NUM
ge	general	ADJ; ADV; PREP
hyph	hyphenated	NUM
imp	imperative	VB
indef	indefinite	ART
infin	infinitive	VB
ingp	<i>-ing</i> participle	ADJ; NADJ; VB
inter	interrogative	PRON
interjec	interjection	MISC
intr	intransitive	VB
lex	lexical	VB
modal	modal	VB
montr	monotransitive	VB
mult	multiplicative	NUM
neg	negative	ADV; PRON; VB
nomplu	plural nominal	ADJ
nomposs	nominal possessive	PRON
nonass	non-assertive	PRON
number	number	N; PRON
obrack	opening bracket	PUNC
one	<i>one</i>	PROFM; PRON
oquo	opening quote	PUNC
ord	ordinal	ADJ NUM
other	other	PUNC
pass	passive voice	VB
past	past tense	VB
per	period	PUNC
perf	perfective aspect	VB
pers	personal	PRON
phras	phrasal	ADV; PREP
phrase	phrase	PROFM
plu	plural	N; NUM; PROFM; PRON
pos	positive	ADJ; ADV; NADJ
poss	possessive	PRON
prefix	prefix	MISC
pres	present tense	VB
procl	proclitic	PRON; VB
prog	progressive aspect	VB
qm	question mark	PUNC
quant	quantifying	PRON
recip	reciprocal	PRON
rel	relative	PRON
scolon	semi colon	PUNC
self	<i>-self / -selves</i>	PRON
semi	semi	VB
semip	semi followed by <i>-ing</i> participle	VB
sing	singular	N; NUM; PROFM; PRON
so	<i>so</i>	PROFM
subjun	subjunctive	VB
subord	subordinating	CONJUNC
such	<i>such</i>	PRON
suffix	suffix	MISC
sup	superlative	ADJ; ADV; NADJ
to	<i>to</i>	PRTCL

univ  
wh  
with

universal  
*wh-*  
*with*

PRON  
ADV  
PRTCL

## APPENDIX 2: INVENTORY OF TOSCA-ICLE TAGS

### Possible label combinations (220)

The following list of possible tags does not contain ditto tags. In principle all multi-token units can be discontinuous. For this reason the feature 'disc' can occur in all ditto tags. As ditto tags do not occur in the list, tags with the discontinuity feature are absent, too.

ADJ(ge,comp)	NUM(frac,sing)	PRON(such)
ADJ(ge,pos,edp)	NUM(hyph,plu)	PRON(univ)
ADJ(ge,pos,ingp)	NUM(hyph,sing)	
ADJ(ge,pos)	NUM(mult)	PRTCL(for)
ADJ(ge,sup)	NUM(ord,plu)	PRTCL(to)
ADJ(ord)	NUM(ord,sing)	PRTCL(with)
ADJ(ord,nomplu)		
	PREP(ge)	PUNC(cbrack)
ADV(connec)	PREP(phras)	PUNC(colon)
ADV(ge,pos)		PUNC(comma)
ADV(ge,comp)	PROFM(conj)	PUNC(cquo)
ADV(ge,sup)	PROFM(one,sing)	PUNC(dash)
ADV(neg)	PROFM(one,plu)	PUNC(ellip)
ADV(phras)	PROFM(so,phrase)	PUNC(exm)
ADV(wh)	PROFM(so,clause)	PUNC(obrack)
		PUNC(oquo)
ART(def)	PRON(antit)	PUNC(other)
ART(indef)	PRON(antit,procl)	PUNC(per)
	PRON(ass)	PUNC(qm)
CONJUNC(coord)	PRON(cleft)	PUNC(scolon)
CONJUNC(subord)	PRON(cleft,procl)	
	PRON(dem,number)	VB(aux,do,imp)
EXTHERE	PRON(dem,plu)	VB(aux,do,imp,neg)
	PRON(dem,sing)	VB(aux,do,past)
GENM	PRON(exclam)	VB(aux,do,past,neg)
	PRON(inter)	VB(aux,do,pres)
MISC(discourse)	PRON(inter,poss)	VB(aux,do,pres,encl)
MISC(foreign)	PRON(neg)	VB(aux,do,pres,neg)
MISC(interjec)	PRON(nonass)	VB(aux,do,pres,procl)
MISC(suffix)	PRON(nomposs,number)	VB(aux,modal,edp)
MISC(prefix)	PRON(nomposs,plu)	VB(aux,modal,infin)
	PRON(nomposs,sing)	VB(aux,modal,ingp)
N(plu,collect)	PRON(one)	VB(aux,modal,past)
N(sing,collect)	PRON(pers,number)	VB(aux,modal,past,encl)
N(number)	PRON(pers,plu)	VB(aux,modal,past,neg)
N(plu)	PRON(pers,plu,encl)	VB(aux,modal,pres)
N(sing)	PRON(pers,sing)	VB(aux,modal,pres,encl)
	PRON(pers,sing,procl)	VB(aux,modal,pres,neg)
NADJ(comp)	PRON(poss,number)	VB(aux,modal,subjunct)
NADJ(pos,edp)	PRON(poss,plu)	VB(aux,modal,subjunct,neg)
NADJ(pos,ingp)	PRON(poss,sing)	VB(aux,pass,edp)
NADJ(pos)	PRON(quant)	VB(aux,pass,imp)
NADJ(sup)	PRON(recip)	VB(aux,pass,infin)
	PRON(rel)	VB(aux,pass,ingp)
NUM(card,plu)	PRON(rel,poss)	VB(aux,pass,past)
NUM(card,sing)	PRON(self,plu)	VB(aux,pass,past,neg)
NUM(frac,plu)	PRON(self,sing)	VB(aux,pass,pres)

VB(aux,pass,pres,encl)	VB(aux,semi,pres,encl)	VB(lex,ditr,past)
VB(aux,pass,pres,neg)	VB(aux,semi,pres,neg)	VB(lex,ditr,pres)
VB(aux,pass,subjun)	VB(aux,semi,subjun)	VB(lex,ditr,subjun)
VB(aux,pass,subjun,neg)	VB(aux,semip,edp)	VB(lex,dimontr,edp)
VB(aux,perf,edp)	VB(aux,semip,imp)	VB(lex,dimontr,imp)
VB(aux,perf,infin)	VB(aux,semip,infin)	VB(lex,dimontr,infin)
VB(aux,perf,infin,encl)	VB(aux,semip,ingp)	VB(lex,dimontr,ingp)
VB(aux,perf,ingp)	VB(aux,semip,past)	VB(lex,dimontr,past)
VB(aux,perf,past)	VB(aux,semip,pres)	VB(lex,dimontr,pres)
VB(aux,perf,past,encl)	VB(aux,semip,subjun)	VB(lex,dimontr,subjun)
VB(aux,perf,past,neg)	VB(lex,cop,edp)	VB(lex,intr,edp)
VB(aux,perf,pres)	VB(lex,cop,imp)	VB(lex,intr,imp)
VB(aux,perf,pres,encl)	VB(lex,cop,infin)	VB(lex,intr,infin)
VB(aux,perf,pres,neg)	VB(lex,cop,ingp)	VB(lex,intr,ingp)
VB(aux,perf,subjun)	VB(lex,cop,past)	VB(lex,intr,past)
VB(aux,prog,edp)	VB(lex,cop,past,neg)	VB(lex,intr,past,neg)
VB(aux,prog,infin)	VB(lex,cop,pres)	VB(lex,intr,pres)
VB(aux,prog,past)	VB(lex,cop,pres,encl)	VB(lex,intr,pres,encl)
VB(aux,prog,past,neg)	VB(lex,cop,pres,neg)	VB(lex,intr,pres,neg)
VB(aux,prog,pres)	VB(lex,cop,subjun)	VB(lex,intr,subjun)
VB(aux,prog,pres,encl)	VB(lex,cop,subjun,neg)	VB(lex,montr,edp)
VB(aux,prog,pres,neg)	VB(lex,cxtr,edp)	VB(lex,montr,imp)
VB(aux,prog,subjun)	VB(lex,cxtr,imp)	VB(lex,montr,infin)
VB(aux,prog,subjun,neg)	VB(lex,cxtr,infin)	VB(lex,montr,ingp)
VB(aux,semi,edp)	VB(lex,cxtr,ingp)	VB(lex,montr,past)
VB(aux,semi,imp)	VB(lex,cxtr,past)	VB(lex,montr,past,encl)
VB(aux,semi,infin)	VB(lex,cxtr,pres)	VB(lex,montr,past,neg)
VB(aux,semi,ingp)	VB(lex,cxtr,subjun)	VB(lex,montr,pres)
VB(aux,semi,past)	VB(lex,ditr,edp)	VB(lex,montr,pres,encl)
VB(aux,semi,past,neg)	VB(lex,ditr,imp)	VB(lex,montr,pres,neg)
VB(aux,semi,pres)	VB(lex,ditr,infin)	VB(lex,montr,subjun)
VB(aux,semi,pres,ellipt)	VB(lex,ditr,ingp)	

### Tags for extra-textual material (2)

MARKUP  
UNTAG