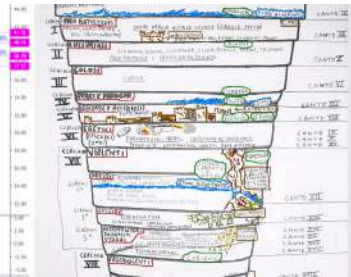
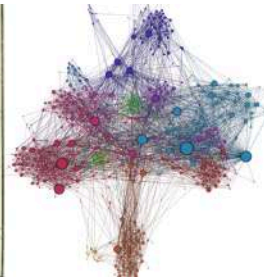


```
# for details, see https://review.docs.microsoft.com/en-us/visualstudio/ctvs/sql-server
# Test code
library(RODBC)
channel <- odbcDriverConnect(dbConnection)
InputDataSet <- sqlQuery(channel, iconv(paste(readlines(
  'c:/proj/rproject1/rproject1/storedprocedure.query.sql',
  encoding = 'UTF-8', warn = FALSE), collapse = '\n'), from = 'UTF-8',
  to = 'ASCII', sub = ''))
odbcClose(channel)
OutputDataSet <- InputDataSet
```



# TreeTagger

Giovanni Pietro Vitali – University College Cork

[giovannipetrovitali@gmail.com](mailto:giovannipetrovitali@gmail.com)

<https://github.com/digitalkoine>

<https://ucc-ie.academia.edu/GiovanniPietroVitali>



**ucc**

Coláiste na hOllscoile Corcaigh, Éire  
University College Cork, Ireland



# TreeTagger

The TreeTagger is a tool for annotating text with part-of-speech and lemma information. It was developed by Helmut Schmid in the TC project at the Institute for Computational Linguistics of the University of Stuttgart. The TreeTagger has been successfully used to tag German, English, French, Italian, Danish, Swedish, Norwegian, Dutch, Spanish, Bulgarian, Russian, Portuguese, Galician, Greek, Chinese, Swahili, Slovak, Slovenian, Latin, Estonian, Polish, Romanian, Czech, Coptic and old French texts and is adaptable to other languages if a lexicon and a manually tagged training corpus are available.



Source: <https://www.cis.uni-muenchen.de/~schmid/tools/TreeTagger/>

# Links

The TreeTagger can also be used as a chunker for English, German, French, and Spanish.

The tagger is described in the following two papers:

- Helmut Schmid (1995): [Improvements in Part-of-Speech Tagging with an Application to German](#). *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.
- Helmut Schmid (1994): [Probabilistic Part-of-Speech Tagging Using Decision Trees](#). *Proceedings of International Conference on New Methods in Language Processing*, Manchester, UK.



# Links to go

- [Graphical Interface](#) for the Windows version of the TreeTagger (developed by Ciarán Ó Duibhín)
- [Serge Sharoffs web page](#) where you can download a tokenizer and a parameter file for Chinese.
- [Pablo Gamallos web page](#) who provides resources for Portuguese and Galician.
- [Achim Stein's web page](#) on French and old French POS tagging with the TreeTagger
- [Python Wrapper](#) for the TreeTagger (developed by Laurent Pointal)
- [Java Wrapper](#) for the TreeTagger (developed by Richard Eckart de Castilho)
- [JavaScript wrapper](#) for the TreeTagger (developed by Thomas Holloway)
- [Perl module](#) for calling the TreeTagger and manipulating its output (developed by Aris Xanthos)
- [R wrapper](#) for the TreeTagger (developed by Meik Michalke)
- [Ruby wrapper](#) for the TreeTagger (developed by Andrei Beliankou)
- [UIMA wrapper](#) for the TreeTagger (developed at the UKP Lab)
- Giuseppe Attardi's [online interface](#) to the TreeTagger.
- [Italian Online Tagger](#) at the University for Foreigners Perugia
- Another [Online Tagger](#) created by Carlos Eduardo
- Webpage on the translation of German texts into [easy-to-read language](#) (in German)

# English Tagset Penn Treebank Project

Tag Number	English TagSet	Penn Treebank Project
1	CC	Coordinating conjunction
2	CD	Cardinal number
3	DT	Determiner
4	EX	Existential there
5	FW	Foreign word
6	IN	Preposition or subordinating conjunction
7	JJ	Adjective
8	JJR	Adjective, comparative
9	JJS	Adjective, superlative
10	LS	List item marker
11	MD	Modal
12	NN	Noun, singular or mass
13	NNS	Noun, plural
14	NP	Proper noun, singular
15	NPS	Proper noun, plural
16	PDT	Predeterminer
17	POS	Possessive ending
18	PP	Personal pronoun
19	PP\$	Possessive pronoun
20	RB	Adverb
21	RBR	Adverb, comparative
22	RBS	Adverb, superlative
23	RP	Particle
24	SYM	Symbol
25	TO	to
26	UH	Interjection
27	VB	Verb, base form
28	VBD	Verb, past tense
29	VBG	Verb, gerund or present participle
30	VBN	Verb, past participle
31	VBP	Verb, non-3rd person singular present
32	VBZ	Verb, 3rd person singular present
33	WDT	Wh-determiner
34	WP	Wh-pronoun
35	WP\$	Possessive wh-pronoun
36	WRB	Wh-adverb

# English Tagset - The BNC Basic (C5) Tagset

Tag Number	English TagSet	The BNC Basic (C5) Tagset
1	AJO	Adjective (general or positive) (e.g. good, old, beautiful)
2	AJC	Comparative adjective (e.g. better, older)
3	AJS	Superlative adjective (e.g. best, oldest)
4	ATO	Article (e.g. the, a, an, no) [N.B. no is included among articles, which are defined here as determiner words which typically begin a noun phrase, but which cannot occur as the head of a noun phrase.]
5	AVO	General adverb: an adverb not subclassified as AVP or AVQ (see below) (e.g. often, well, longer (adv.), furthest. [Note that adverbs, unlike adjectives, are not tagged as positive, comparative, or superlative.This is because of the relative rarity of comparative and superlative adverbs.]
6	AVP	Adverb particle (e.g. up, off, out) [N.B. AVP is used for such "prepositional adverbs", whether or not they are used idiomatically in a phrasal verb: e.g. in 'Come out here' and 'I can't hold out any longer', the same AVP tag is used for out.
7	AVQ	Wh-adverb (e.g. when, where, how, why, wherever) [The same tag is used, whether the word occurs in interrogative or relative use.]
8	CJC	Coordinating conjunction (e.g. and, or, but)
9	CJS	Subordinating conjunction (e.g. although, when)
10	CJT	The subordinating conjunction that [N.B. that is tagged CJT when it introduces not only a nominal clause, but also a relative clause, as in 'the day that follows Christmas'. Some theories treat that here as a relative pronoun, whereas others treat it as a conjunction.We have adopted the latter analysis.]
11	CRD	Cardinal number (e.g. one, 3, fifty-five, 3609)
12	DPS	Possessive determiner (e.g. your, their, his)
13	DT0	General determiner: i.e. a determiner which is not a DTQ. [Here a determiner is defined as a word which typically occurs either as the first word in a noun phrase, or as the head of a noun phrase. E.g. This is tagged DT0 both in 'This is my house' and in 'This house is mine'.]
14	DTQ	Wh-determiner (e.g. which, what, whose, whichever) [The category of determiner here is defined as for DT0 above. These words are tagged as wh-determiners whether they occur in interrogative use or in relative use.]
15	EXO	Existential there, i.e. there occurring in the there is ... or there are ... construction
16	ITJ	Interjection or other isolate (e.g. oh, yes, mhm, wow)
17	NN0	Common noun, neutral for number (e.g. aircraft, data, committee) [N.B. Singular collective nouns such as committee and team are tagged NN0, on the grounds that they are capable of taking singular or plural agreement with the following verb: e.g. 'The committee disagrees/disagrees'.]
18	NN1	Singular common noun (e.g. pencil, goose, time, revelation)
19	NN2	Plural common noun (e.g. pencils, geese, times, revelations)
20	NP0	Proper noun (e.g. London, Michael, Mars, IBM) [N.B. the distinction between singular and plural proper nouns is not indicated in the tagset, plural proper nouns being a comparative rarity.]
21	ORD	Ordinal numeral (e.g. first, sixth, 77th, last) . [N.B. The ORD tag is used whether these words are used in a nominal or in an adverbial role. Next and last, as "general ordinals", are also assigned to this category.]
22	PNI	Indefinite pronoun (e.g. none, everything, one [as pronoun], nobody) [N.B. This tag applies to words which always function as [heads of] noun phrases. Words like some and these, which can also occur before a noun head in an article-like function, are tagged as determiners (see DT0 and ATO above).]
23	PNP	Personal pronoun (e.g. I, you, them, ours) [Note that possessive pronouns like ours and theirs are tagged as personal pronouns.]
24	PNQ	Wh-pronoun (e.g. who, whoever, whom) [N.B. These words are tagged as wh-pronouns whether they occur in interrogative or in relative use.]
25	PNX	Reflexive pronoun (e.g. myself, yourself, itself, ourselves)
26	POS	The possessive or genitive marker 's or ' (e.g. for 'Peter's or somebody else's', the sequence of tags is: NP0 POS CJC PNI AV0 POS)
27	PRF	The preposition of. Because of its frequency and its almost exclusively postnominal function, of is assigned a special tag of its own.
28	PRP	Preposition (except for of) (e.g. about, at, in, on, on behalf of, with)
29	PUL	Punctuation: left bracket
30	PUN	Punctuation: general separating mark
31	PUQ	Punctuation: quotation mark
32	PUR	Punctuation: right bracket
33	TO0	Infinitive marker to
34	UNC	Unclassified items which are not appropriately classified as items of the English lexicon. [Items tagged UNC include foreign (non-English) words, special typographical symbols, formulae, and (in spoken language) hesitation fillers such as er and erm.]
35	VBB	The present tense forms of the verb BE, except for is, 's: i.e. am, are, 'm, 're and be [subjunctive or imperative]
36	VBD	The past tense forms of the verb BE: was and were
37	VBG	The -ing form of the verb BE: being
38	VBI	The infinitive form of the verb BE: be
39	VBN	The past participle form of the verb BE: been
40	VBZ	The -s form of the verb BE: is, 's
41	VDB	The finite base form of the verb BE: do
42	VDD	The past tense form of the verb DO: did
43	VDG	The -ing form of the verb DO: doing
44	VDI	The infinitive form of the verb DO: do
45	VDN	The past participle form of the verb DO: done
46	VDZ	The -s form of the verb DO: does, 's
47	VHB	The finite base form of the verb HAVE: have, 've
48	VHD	The past tense form of the verb HAVE: had, 'd
49	VHG	The -ing form of the verb HAVE: having
50	VHI	The infinitive form of the verb HAVE: have
51	VHN	The past participle form of the verb HAVE: had
52	VHZ	The -s form of the verb HAVE: has, 's
53	VM0	Modal auxiliary verb (e.g. will, would, can, could, 'll, 'd)
54	VVB	The finite base form of lexical verbs (e.g. forget, send, live, return) [including the imperative and present subjunctive]
55	VVD	The past tense form of lexical verbs (e.g. forgot, sent, lived, returned)
56	VVG	The -ing form of lexical verbs (e.g. forgetting, sending, living, returning)
57	VVI	The infinitive form of lexical verbs (e.g. forget, send, live, return)
58	VVN	The past participle form of lexical verbs (e.g. forgotten, sent, lived, returned)
59	VVZ	The -s form of lexical verbs (e.g. forgets, sends, lives, returns)
60	XX0	The negative particle not or n't
61	ZZ0	Alphabetical symbols (e.g. A, a, B, b, c, d)

# Spanish Tagset

Tag Number	Spanish TagSet	
1	ACRNM	acronym (ISO, CEI)
2	ADJ	Adjectives (mayores, mayor)
3	ADV	Adverbs (muy, demasiado, cómo)
4	ALFP	Plural letter of the alphabet (As/Aes, bes)
5	ALFS	Singular letter of the alphabet (A, b)
6	ART	Articles (un, las, la, unas)
7	BACKSLASH	backslash (\)
8	CARD	Cardinals
9	CC	Coordinating conjunction (y, o)
10	CCAD	Adversative coordinating conjunction (pero)
11	CCNEG	Negative coordinating conjunction (ni)
12	CM	comma (,)
13	CODE	Alphanumeric code
14	COLON	colon (:)
15	CQUE	que (as conjunction)
16	CSUBF	Subordinating conjunction that introduces finite clauses (apenas)
17	CSUBI	Subordinating conjunction that introduces infinite clauses (al)
18	CSUBX	Subordinating conjunction underspecified for subord-type (aunque)
19	DASH	dash (-)
20	DM	Demonstrative pronouns (ésas, ése, esta)
21	DOTS	POS tag for "..."
22	FO	Formula
23	FS	Full stop punctuation marks
24	INT	Interrogative pronouns (quiénes, cuántas, cuánto)
25	ITJN	Interjection (oh, ja)
26	LP	left parenthesis ("(", "[")
27	NC	Common nouns (mesas, mesa, libro, ordenador)
28	NEG	Negation
29	NMEA	measure noun (metros, litros)
30	NMON	month name
31	NP	Proper nouns
32	ORD	Ordinals (primer, primeras, primera)
33	PAL	Portmanteau word formed by a and el
34	PDEL	Portmanteau word formed by de and el
35	PE	Foreign word
36	PERCT	percent sign (%)
37	PNC	Unclassified word
38	PPC	Cillic personal pronoun (le, les)
39	PPO	Possessive pronouns (mi, su, sus)
40	PPX	Cillics and personal pronouns (nos, me, nosotras, te, si)
41	PREP	Negative preposition (sin)
42	PREP	Preposition
43	PREP/DEL	Complex preposition "después del"
44	QT	quotation symbol ("")
45	QU	Quantifiers (sencas, cada)
46	REL	Relative pronouns (cuyas, cuyo)
47	RP	right parenthesis (")", "}")
48	SE	Se (as particle)
49	SEMICOLON	semicolon (;)
50	SLASH	slash (/)
51	SYM	Symbols
52	UMMX	measure unit (MHz, km, mA)
53	VCLlger	cillic gerund verb
54	VCLlnt	cillic infinitive verb
55	VCLlnt	cillic finite verb
56	VEadj	Verb estar, Finite
57	VEfin	Verb estar, Finite
58	VEger	Verb estar, Gerund
59	VEint	Verb estar, Infinitive
60	VHadj	Verb haber, Past participle
61	VHfin	Verb haber, Finite
62	VHger	Verb haber, Gerund
63	VHinf	Verb haber, Infinitive
64	VLadj	Lexical verb, Past participle
65	VLfin	Lexical verb, Finite
66	VLger	Lexical verb, Gerund
67	VLint	Lexical verb, Infinitive
68	VMadj	Modal verb, Past participle
69	VMfin	Modal verb, Finite
70	VMger	Modal verb, Gerund
71	VMint	Modal verb, Infinitive
72	VSadj	Verb ser, Past participle
73	VSfin	Verb ser, Finite
74	VSger	Verb ser, Gerund
75	VSint	Verb ser, Infinitive

# French Tagset

## Achim Stein,

### April 2003

Tag Number	French TagSet	Achim Stein, April 2003
1	ABR	abbreviation
2	ADJ	adjective
3	ADV	adverb
4	DET:ART	article
5	DET:POS	possessive pronoun (ma, ta, ...)
6	INT	interjection
7	KON	conjunction
8	NAM	proper name
9	NOM	noun
10	NUM	numeral
11	PRO	pronoun
12	PRO:DEM	demonstrative pronoun
13	PRO:IND	indefinite pronoun
14	PRO:PER	personal pronoun
15	PRO:POS	possessive pronoun (mien, tien, ...)
16	PRO:REL	relative pronoun
17	PRP	preposition
18	PRP:det	preposition plus article (au, du, aux, des)
19	PUN	punctuation
20	PUN:cit	punctuation citation
21	SENT	sentence tag
22	SYM	symbol
23	VER:cond	verb conditional
24	VER:futu	verb futur
25	VER:impe	verb imperative
26	VER:impf	verb imperfect
27	VER:infi	verb infinitive
28	VER:pper	verb past participle
29	VER:ppre	verb present participle
30	VER:pres	verb present
31	VER:simp	verb simple past
32	VER:subi	verb subjunctive imperfect
33	VER:subp	verb subjunctive present

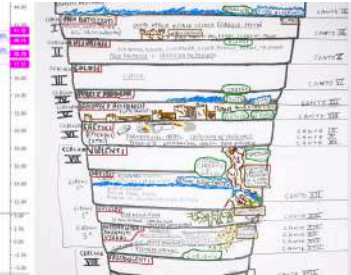
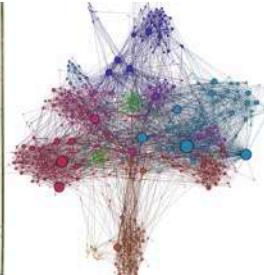


# Italian Tagset

Copyright Prof. Achim Stein, University of Stuttgart

Tag Number	Italian TagSet	Copyright Prof. Achim Stein, University of Stuttgart
1	ABR	abbreviation
2	ADJ	adjective
3	ADV	adverb
4	CON	conjunction
5	DET:def	definite article
6	DET:indef	indefinite article
7	FW	foreign word
8	INT	interjection
9	LS	list symbol
10	NOM	noun
11	NPR	name
12	NUM	numeral
13	PON	punctuation
14	PRE	preposition
15	PRE:det	preposition+article
16	PRO	pronoun
17	PRO:demo	demonstrative pronoun
18	PRO:indef	indefinite pronoun
19	PRO:inter	interrogative pronoun
20	PRO:pers	personal pronoun
21	PRO:poss	possessive pronoun
22	PRO:refl	reflexive pronoun
23	PRO:rela	relative pronoun
24	SENT	sentence marker
25	SYM	symbol
26	VER:cimp	verb conjunctive imperfect
27	VER:cond	verb conditional
28	VER:cpre	verb conjunctive present
29	VER:futu	verb future tense
30	VER:geru	verb gerund
31	VER:impe	verb imperative
32	VER:impf	verb imperfect
33	VER:infi	verb infinitive
34	VER:pper	verb participle perfect
35	VER:ppe	verb participle present
36	VER:pres	verb present
37	VER:refl:infi	verb reflexive infinitive
38	VER:remo	verb simple past

```
# for details, see https://review.docs.microsoft.com/en-us/visualstudio/ctvs/sql-server
# Test code
library(RODBC)
channel <- odbcDriverConnect(dbConnection)
InputDataSet <- sqlQuery(channel, iconv(paste(readlines(
  'c:/proj/rproject1/rproject1/storedprocedure.query.sql',
  encoding = 'UTF-8', warn = FALSE), collapse = '\n'), from = 'UTF-8',
  to = 'ASCII', sub = ''))
odbcClose(channel)
OutputDataSet <- InputDataSet
```



# Distant Reading

# \_end

Giovanni Pietro Vitali – University College Cork

[giovannipetrovitali@gmail.com](mailto:giovannipetrovitali@gmail.com)

<https://github.com/digitalkoine>

<https://ucc-ie.academia.edu/GiovanniPietroVitali>



**ucc**

Coláiste na hOllscoile Corcaigh, Éire  
University College Cork, Ireland

