

## Appendix B1

## POS Tag Set Used with the Brill Tagger

N	Tag	Definition	Examples
1	CC	coordinating conjunction	and, but, nor, or, so, both, either, neither, therefore, etc.
2	CD	cardinal, numeral	1998, nine-thirty, one-tenth, ten million, 0.5, twenty, '98, two, 30-degrees, etc.
3	DT	determiner	a, an, the, this, that, another, any, both, each, every, some, these, those, all, etc.
4	EX	existential THERE	there
5	FW	foreign word	<i>gemeinschaft, vous, je, oui</i> , etc.
6	IN	preposition and subordinate conjunction	for, of, in, with, that, while, despite, beside, although, until, unless, (if, whether), etc.
7	JJ	adjective and adjective	first, regrettable, participatory, oiled, well-deserved, post-
8	JJR	adjective, comparative	cheaper, cleaner, closer, colder, commoner, deadlier, etc.
9	JJS	superlative adjective	cheapest, cleanest, closest, coldest, commonest, deadliest,
10	LS	list item marker	a, b, c, first, second, one, two, CA-0010, etc.
11	MD	modal auxiliary	can, could, may, might, must, shall, should, will, would., ought, dare (=> <i>See Note 1</i> )
12	NN	common noun, singular or mass	letter, desk, investment, communication; VP, CEO (= <i>singular abbreviated noun phrase</i> ), etc.
13	NNP	proper noun, singular	John, Yamada, Japan, etc.
14	NNPS	proper noun, plural	Americans, Yamadas, etc.
15	NNS	common noun, plural	letters, desks, investments, communications; VPs, CEOs (= <i>plural abbreviated noun phrase</i> ), etc.
16	PDT	pre-determiner	all, both, half, many, quite, such, sure, this (e.g. <i>all</i> the time; make <i>this</i> a model), etc.
17	POS	genitive marker	's (as in <i>teacher's, Yamada's</i> , etc.) (=> <i>See Note 2</i> )
18	PRP	personal and/or non-personal pronoun	I, we, he, she, his, her, they, them, it, one, himself, ourselves, etc.
19	PRP\$	possessive pronoun	her, his, mine, my, our, ours, their, your, etc.
20	RB	adverb	occasionally, immediately, surely, lately, etc.
21	RBR	comparative adverb	further, harder, healthier, higher, later, longer, more, etc.
22	RBS	superlative adverb	best, biggest, earliest, first, furthest, hardest, largest, most,
23	RP	particle	aboard, across, along, apart, around, aside, away, back,
24	SYM	symbol	% & ' " " .) . * + , . < = > @ \ , etc. (=> <i>See Note 3</i> )
25	TO	infinitive marker	to (followed by VB) (=> <i>See Note 4</i> )
26	UH	interjection	Goodbye, Goody, Gosh, Wow, Oops, huh, uh, dammit, etc.
27	VB	verb, present tense or bare form	ask, assemble, assign, assume, avoid, begin, buy, believe, etc.
28	VBD	verb, past tense	asked, assembled, assigned, assumed, avoided, began,
29	VBG	verb, present participle or gerund	asking, assembling, assigning, assuming, avoiding, buying, beginning, etc.
30	VBN	verb, past participle	asked, assumed, believe, begun, gone, written, etc.
31	VBP	verb, present, non-third person singular	appear, obtain, comprise, attract, emphasize, postpone, return, etc.
32	VBZ	verb, present, third person singular	appears, obtains, comprises, attracts, emphasizes, postpones, returns, etc.
33	WDT	wh-determiner	that, what, whatever, which, whichever, whatsoever,
34	WP	wh-pronoun	that, what, whatever, whatsoever, which, who, whom, whosoever, whoever, etc.

N	Tag	Definition	Examples
35	WP\$	wh-pronoun, possessive	whose
36	WRB	wh-adverb	how, however, whence, whenever, where, whereby, wherever, wherein, whereof, why

*The following tags have been added to the original tag set of the Brill Tagger.*

37	BE	be, present and past	be, is, am, was, were
38	BEG	be, present participle or gerund	being
39	BEN	be, past participle	been
40	DO	do, present and past	do, does, did
41	DOG	do, present participle or gerund	doing
42	DON	do, past participle	done
43	HV	have, present and past	have, has, had
44	HVG	have, present participle or gerund	having
45	IF	coordinating conjunction, conditional marker	if
46	WZH	coordinating conjunction, conditional marker	whether
47	NEG	negative ending (= n't, 't)	n't, 't (=> See Note 5)

Notes:

- 1) Semi-modals "need" and "have to" are not included.
- 2) A string like "teacher's," therefore, is counted as two words. *e.g.* teacher\_NN 's\_POS
- 3) In the current study, all symbols are assigned with the same symbol connected by the underbar. *e.g.* ,\_ / \_/ :- :%\_%, etc.
- 4) The Brill Tagger in its original form (Ver. 1.14: 1994) does not distinguish the prepositional "to" (to + N) from the infinitival "to" (to + VB). The current version is so modified that it assigns the "TO" tag to the infinitival "to" and the "IN" tag to the prepositional "to."
- 5) A string like "can't," therefore, is counted as two words. *e.g.* can\_MD 't\_NEG

Source: *The University of Pennsylvania (Penn) Treebank Standard Tagset*  
 [Online] URL: <http://www.scs.leeds.ac.uk/ccalas/tagset/penn.html>