

Appendix A: CLAWS Tag Set

The following is the CLAWS tag set from Garside (1997) that was used for the word-class identification.

AJO	Adjective (general or positive) (e.g. <i>good, old, beautiful</i>)	CJC	Coordinating conjunction (e.g. <i>and, or, but</i>)
AJC	Comparative adjective (e.g. <i>better, older</i>)	CJS	Subordinating conjunction (e.g. <i>although, when</i>)
AJS	Superlative adjective (e.g. <i>best, oldest</i>)	CJT	The subordinating conjunction <i>that</i> [N.B. <i>that</i> is tagged CJT when it introduces not only a nominal clause, but also a relative clause, as in 'the day <i>that</i> follows Christmas'. Some theories treat <i>that</i> here as a relative pronoun, whereas others treat it as a conjunction. We have adopted the latter analysis.]
ATO	Article (e.g. <i>the, a, an, no</i>) [N.B. <i>no</i> 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.]	CRD	Cardinal number (e.g. <i>one, 3, fifty-five, 3609</i>)
AVO	General adverb: an adverb not subclassified as AVP or AVQ (see below) (e.g. <i>often, well, longer (adv.), furthest</i> . [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.]	DPS	Possessive determiner (e.g. <i>your, their, his</i>)
AVP	Adverb particle (e.g. <i>up, off, out</i>) [N.B. AVP is used for such "prepositional adverbs", whether or not they are used idiomatically in a phrasal verb: e.g. in 'Come <i>out</i> here' and 'I can't hold <i>out</i> any longer', the same AVP tag is used for <i>out</i> .]	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. <i>This</i> is tagged DT0 both in ' <i>This</i> is my house' and in ' <i>This</i> house is mine'.]
AVQ	<i>Wh</i> -adverb (e.g. <i>when, where, how, why, wherever</i>) [The same tag is used, whether the word occurs in interrogative or relative use.]	DTQ	<i>Wh</i> -determiner (e.g. <i>which, what, whose, whichever</i>) [The category of determiner here is defined as for DT0 above. These words are tagged as <i>wh</i> -determiners whether they occur in interrogative use or in relative use.]
		ITJ	Interjection or other isolate (e.g. <i>oh, yes, mhm, wow</i>)
		NN0	Common noun, neutral for number (e.g. <i>aircraft, data, committee</i>) [N.B. Singular collective nouns such as <i>committee</i> and <i>team</i> are tagged NN0, on the grounds that they are capable of taking singular or plural

	agreement with the following verb: e.g. 'The <i>committee</i> disagrees/disagree'.]		pronouns whether they occur in interrogative or in relative use.]
NN1	Singular common noun (e.g. <i>pencil, goose, time, revelation</i>)	PNX	Reflexive pronoun (e.g. <i>myself, yourself, itself, ourselves</i>)
NN2	Plural common noun (e.g. <i>pencils, geese, times, revelations</i>)	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)
NP0	Proper noun (e.g. <i>London, Michael, Mars, IBM</i>) [N.B. the distinction between singular and plural proper nouns is not indicated in the tagset, plural proper nouns being a comparative rarity.]	PRF	The preposition <i>of</i> . Because of its frequency and its almost exclusively postnominal function, <i>of</i> is assigned a special tag of its own.
ORD	Ordinal numeral (e.g. <i>first, sixth, 77th, last</i>). [N.B. The ORD tag is used whether these words are used in a nominal or in an adverbial role. <i>Next</i> and <i>last</i> , as "general ordinals", are also assigned to this category.]	PRP	Preposition (except for <i>of</i>) (e.g. <i>about, at, in, on, on behalf of, with</i>)
PNI	Indefinite pronoun (e.g. <i>none, everything, one</i> [as pronoun], <i>nobody</i>) [N.B. This tag applies to words which always function as [heads of] noun phrases. Words like <i>some</i> and <i>these</i> , which can also occur before a noun head in an article-like function, are tagged as determiners (see DT0 and AT0 above).]	TOO	Infinitive marker <i>to</i>
PNP	Personal pronoun (e.g. <i>I, you, them, ours</i>) [Note that possessive pronouns like <i>ours</i> and <i>theirs</i> are tagged as personal pronouns.]	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 <i>er</i> and <i>erm</i> .]
PNQ	Wh-pronoun (e.g. <i>who, whoever, whom</i>) [N.B. These words are tagged as wh-	VBB	The present tense forms of the verb BE, except for <i>is, 's</i> : i.e. <i>am, are, 'm, 're</i> and <i>be</i> [subjunctive or imperative]
		VBD	The past tense forms of the verb BE: <i>was</i> and <i>were</i>
		VBG	The <i>-ing</i> form of the verb BE: <i>being</i>
		VBI	The infinitive form of the verb BE: <i>be</i>
		VBN	The past participle form of the verb BE: <i>been</i>
		VBZ	The <i>-s</i> form of the verb BE: <i>is, 's</i>
		VDB	The finite base form of the verb BE: <i>do</i>
		VDD	The past tense form of the verb DO: <i>did</i>
		VDG	The <i>-ing</i> form of the verb DO: <i>doing</i>
		VDI	The infinitive form of the verb DO: <i>do</i>
		VDN	The past participle form of the verb DO: <i>done</i>
		VDZ	The <i>-s</i> form of the verb DO: <i>does, 's</i>

VHB	The finite base form of the verb HAVE: <i>have, 've</i>		the imperative and present subjunctive]
VHD	The past tense form of the verb HAVE: <i>had, 'd</i>	VVD	The past tense form of lexical verbs (e.g. <i>forgot, sent, lived, returned</i>)
VHG	The <i>-ing</i> form of the verb HAVE: <i>having</i>	VVG	The <i>-ing</i> form of lexical verbs (e.g. <i>forgetting, sending, living, returning</i>)
VHI	The infinitive form of the verb HAVE: <i>have</i>	VVI	The infinitive form of lexical verbs (e.g. <i>forget, send, live, return</i>)
VHN	The past participle form of the verb HAVE: <i>had</i>	VVN	The past participle form of lexical verbs (e.g. <i>forgotten, sent, lived, returned</i>)
VHZ	The <i>-s</i> form of the verb HAVE: <i>has, 's</i>	VVZ	The <i>-s</i> form of lexical verbs (e.g. <i>forgets, sends, lives, returns</i>)
VM0	Modal auxiliary verb (e.g. <i>will, would, can, could, 'll, 'd</i>)	XX0	The negative particle <i>not</i> or <i>n't</i>
VVB	The finite base form of lexical verbs (e.g. <i>forget, send, live, return</i>) [Including	ZZ0	Alphabetical symbols (e.g. <i>A, a, B, b, c, d</i>)