

Part of Speech Tags for Assignment 4

The following is a brief description of the Basic Tagset used in the tagging of the whole 100-million-word BNC.

1. A List of Grammatical Tags with Brief Definitions and Clarifications

Each tag consists of three characters. Generally, the first two characters indicate the general part of speech, and the third character is used to indicate a subcategory. When the most general, unmarked category of a part of speech is indicated, in general the third character is 0. (For example, AJ0 is the tag for the most general class of adjectives.)

AJ0 Adjective (general or positive) (e.g. *good, old, beautiful*)

AJC Comparative adjective (e.g. *better, older*)

AJS Superlative adjective (e.g. *best, oldest*)

AT0 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.]

AV0 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.]

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*.

AVQ *Wh*-adverb (e.g. *when, where, how, why, wherever*) [The same tag is used, whether the word occurs in interrogative or relative use.]

CJC Coordinating conjunction (e.g. *and, or, but*)

CJS Subordinating conjunction (e.g. *although, when*)

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.]

CRD Cardinal number (e.g. *one, 3, fifty-five, 3609*)

DPS Possessive determiner (e.g. *your, their, his*)

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'.]

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.]

EX0 Existential *there*, i.e. *there* occurring in the *there is ...* or *there are ...* construction

ITJ Interjection or other isolate (e.g. *oh, yes, mhm, wow*)

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/disagree'.]

NN1 Singular common noun (e.g. *pencil, goose, time, revelation*)

NN2 Plural common noun (e.g. *pencils, geese, times, revelations*)

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.]

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.]

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 AT0 above).]

PNP Personal pronoun (e.g. *I, you, them, ours*) [Note that possessive pronouns like *ours* and *theirs* are tagged as personal pronouns.]

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.]

PNX Reflexive pronoun (e.g. *myself, yourself, itself, ourselves*)

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)

PRF The preposition *of*. Because of its frequency and its almost exclusively postnominal function, *of* is assigned a special tag of its own.

PRP Preposition (except for *of*) (e.g. *about, at, in, on, on behalf of, with*)

PUL Punctuation: left bracket - i.e. (or [

PUN Punctuation: general separating mark - i.e. . , ! , : ; - or ?

PUQ Punctuation: quotation mark - i.e. ' or "

PUR Punctuation: right bracket - i.e.) or]

TOO Infinitive marker *to*

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*.]

VBB The present tense forms of the verb BE, except for *is*, 's: i.e. *am, are, 'm, 're* and *be* [subjunctive or imperative]

VBD The past tense forms of the verb BE: *was* and *were*

VBG The *-ing* form of the verb BE: *being*

VBI The infinitive form of the verb BE: *be*

VBN The past participle form of the verb BE: *been*

VBZ The *-s* form of the verb BE: *is*, 's

VDB The finite base form of the verb BE: *do*

VDD The past tense form of the verb DO: *did*

VDG The *-ing* form of the verb DO: *doing*

VDI The infinitive form of the verb DO: *do*

VDN The past participle form of the verb DO: *done*

VDZ The *-s* form of the verb DO: *does*, 's

VHB The finite base form of the verb HAVE: *have*, 've

VHD The past tense form of the verb HAVE: *had*, 'd

VHG The *-ing* form of the verb HAVE: *having*

VHI The infinitive form of the verb HAVE: *have*

VHN The past participle form of the verb HAVE: *had*

VHZ The *-s* form of the verb HAVE: *has*, 's

VM0 Modal auxiliary verb (e.g. *will, would, can, could, 'll, 'd*)

VVB The finite base form of lexical verbs (e.g. *forget, send, live, return*) [Including the imperative and present subjunctive]

VVD The past tense form of lexical verbs (e.g. *forgot, sent, lived, returned*)

VVG The *-ing* form of lexical verbs (e.g. *forgetting, sending, living, returning*)

VVI The infinitive form of lexical verbs (e.g. *forget, send, live, return*)

VVN The past participle form of lexical verbs (e.g. *forgotten, sent, lived, returned*)

VVZ The -s form of lexical verbs (e.g. *forgets, sends, lives, returns*)

XX0 The negative particle *not* or *n't*

ZZ0 Alphabetical symbols (e.g. *A, a, B, b, c, d*)

Total number of grammatical tags in the BNC Basic Tagset: 61

2. A List of Ambiguity Tags

AJ0-AV0 AJ0-VVN AJ0-VVD AJ0-NN1 AJ0-VVG

AVP-PRP AVQ-CJS CJS-PRP CJT-DT0 CRD-PNI

NN1-NP0 NN1-VVB NN1-VVG NN2-VVZ VVD-VVN