The SPECIALIST LEXICON

Allen C. Browne, Alexa T. McCray, Suresh Srinivasan

Lister Hill National Center for Biomedical Communications
National Library of Medicine
Bethesda, Maryland

REVISED: JUNE 2018
2018 Editors: Amanda Payne, Destinee Tormey

15. Classification Types

1. Introduction

1.1 General Description

A lexicon, recording information specific to individual lexical items, is necessarily a core component of any natural language processing system. The SPECIALIST lexicon has been developed to provide the lexical information needed for the SPECIALIST Natural Language Processing System. It is intended to be a general English lexicon that includes many biomedical terms. Coverage includes both general English words as well as more specialized biomedical vocabulary discovered in the NLM Test Collection and the UMLS Metathesaurus. The lexicon entry for each word or term records the syntactic, morphological, and graphemic information. Syntactic information includes syntactic category (part of speech), and complementation patterns for verbs, adjectives and nouns, as well as positional and modification types for adjectives and adverbs. Inflectional morphology is indicated for those syntactic categories which inflect, and spelling variation is recorded for each lexical item known to exhibit such variation.

The lexicon consists of a set of lexical entries, with one entry for each spelling or set of spelling variants in a particular part of speech. Lexical items may be "multi-word" terms made up of other words if the multi-word term is determined to be a lexical item by its presence as a term in general English or medical dictionaries, or in medical thesauri such as MeSH. Expansions of generally used acronyms and abbreviations are also allowed as multi-word terms.

The lexical entry is a frame structure consisting of slots and fillers. Each entry is enclosed in braces ({...}) and identified by a unique entry number (EUI) recorded as the filler of the **entry=** slot. The EUI is a seven digit number preceded by the letter "E". The **cat=** slot indicates the part of speech of the entry and the **base=** slot indicates the base form of the entry. The base form is the uninflected citation form of the lexical item; the infinitive in the case of a verb; the singular in the case of a noun; and the positive in the case of an inflecting adjective or adverb. Optionally a **spelling_variants=** slot records spelling variants of the base form, including variants used in other English dialects like UK English, Australian English, and Canadian English. Spelling variants are listed in the following order: pure ASCII terms first, then terms without punctuation, then shortest length terms, and alphabetized after that. The spelling variant listed first by this metric (the filler of "base=") is called the citation form, while the other forms are not. That means each entry in the lexicon may have multiple base forms, or spelling variants, but each has just one citation form. The lexical entries for the citation form *anesthetic* given below illustrate some of the features of a SPECIALIST lexical entry:

1.
```
{base=anesthetic
    spelling_variant=anaesthetic
    entry=E0008769
        cat=noun
        variants=reg
    }

{base=anesthetic
    spelling_variant=anaesthetic
    entry=E0008770
        cat=adj
        variants=inv
        position=attrib(3)
    }
```

There are two entries for the base form *anesthetic,* a noun entry and an adjective entry.

The **variants=** slot contains a code indicating the inflectional morphology of each entry; the filler **reg** in the noun entry indicates that the noun *anesthetic* is a count noun which undergoes regular English plural formation (*anesthetics*); **inv** in the **variants=** slot of the adjective entry indicates that the adjective *anesthetic* does not form a comparative or superlative. The **position=** slot indicates that the adjective *anesthetic* is attributive and appears after color adjectives in the normal adjective order.

Lexical entries are not currently divided into senses. So, an entry represents a spelling-category pairing regardless of semantics. The noun *act* has two senses, both of which show a capitalized and lowercase spelling; an act of a play and an act of law. Since both senses share the same spellings and syntactic category, they are represented by a single lexical entry in the current lexicon.

2.
```
{base=Act
    spelling_variant=act
    entry=E0000154
        cat=noun
        variants=reg
    }
```

When different senses have different syntactic behavior, codes for each behavior are recorded in a single entry. For example, *beer* has two senses: the alcoholic beverage and the amount of a standard container of that beverage.

3a. Patients who drank beer recovered more slowly than patients who drank wine.

3b. 56 patients reported drinking more than five beers a day.

The first sense illustrated in 3a. is a mass (uncount) noun. The second sense illustrated in 3b. is a regular (count) noun. In cases like this the appropriate codes for both senses are included in the entry.

4.
```
{base=beer
```

entry=E0012226
cat=noun
variants=uncount
variants=reg
}

Two codes will also appear in cases where the lexical item is both count and uncount without a sense distinction. *Abdominal delivery* denotes the same procedure whether it appears as an uncount noun as in 5a. or a count noun as in 5b.

5a. Abdominal delivery is the procedure of choice in this situation.

5b. Abdominal deliveries are more common these days.

So the lexical record for *abdominal delivery* includes both codes.

6. {base=abdominal delivery
entry=E0006453
cat=noun
variants=uncount
variants=reg
}

Other syntactic codes such as complement codes for verbs, adjectives and nouns are similarly grouped without regard to sense.

1.2 The Development of the Lexicon

Words and terms are selected for lexical coding from a variety of sources. Approximately 20,000 words from the UMLS Test Collection of MEDLINE abstracts together with words which appear both in the UMLS Metathesaurus and *Dorland's Illustrated Medical Dictionary* form the core of the words entered. In addition, an effort has been made to include words from the general English vocabulary. The 10,000 most frequent words listed in The American Heritage Word Frequency Book and the list of 2,000 words used in definitions in Longman's Dictionary of Contemporary English have also been coded. Since the majority of the words selected for coding are nouns, an effort has been made to include verbs and adjectives by identifying verbs in current MEDLINE citation records, by using the *Computer Usable Oxford Advanced Learner's Dictionary*, and by identifying potential adjectives from *Dorland's Illustrated Medical Dictionary* using heuristics developed by McCray and Srinivasan (1990). A systematic model was developed to retrieve words and terms from MEDLINE in 2014. We plan to apply this model to consumer health related data and other resources to enrich the coverage of the Lexicon.

A variety of reference sources was used in coding lexical records. Coding was based on actual usage in the NLM Test Collection, dictionaries of general English, primarily learner's dictionaries which record the kind of syntactic information needed for NLP, and medical dictionaries. *Longman's Dictionary of Contemporary English*, *Dorland's Illustrated Medical Dictionary*, *Collins* COBUILD *Dictionary, The Oxford Advanced Learner's Dictionary*, and *Webster's Medical Desk Dictionary* were make available to the coders. The early development of the lexicon coding scheme relied heavily on the coding scheme used in the first edition of the *Longman Dictionary of Contemporary English*, with only minor deviation from that scheme. But changes in the current lexicon scheme such as the addition of **modification_type** codes for adverbs

and **position** codes for adjectives have moved the SPECIALIST lexicon coding system farther away from Longman's and have required increased use of other lexicographic sources such as *Collins* COBUILD *Dictionary*.

The SPECIALIST LEXICON (unit lexical record formatted file) along with relational files have been released annually as one of the UMLS Knowledge Sources since 1994. Number words, including cardinal, ordinal and fractions, were added to the Lexicon release in 2003. XML format of unit lexical records, XML schemas and JAXB (Java Architecture XML Binding) APIs were available in the LexCheck package released in 2005. The Lexicon migrated to Unicode and has been released in UTF-8 format since 2006. Derivations with negation information (DM.DB) and synonyms (SM.DB) are generated to synchronize with the Lexicon annual release since 2013 and 2017, respectively.

1.3 Verbs

The basic sentence patterns of a language are determined by the number and nature of the complements taken by verbs, since the complementation of the main verb largely determines the structural skeleton of a sentence. SPECIALIST recognizes five broad complementation patterns: intransitive, transitive, ditransitive, linking and complex-transitive. These complementation classes are manifested in the lexicon as slots filled by codes further specifying the verbs' complementation pattern. Table 1 indicates the slot name associated with each complementation class and the page on which that class and its elaborations are discussed.

Complementation Class	Slot Name	Page
intransitive	intran	page 33
transitive	tran=	page 34
ditransitive	ditran=	page 35
linking	link=	page 36
complex-transitive	cplxtran=	page 37

Table 1. Verb Complementation patterns in SPECIALIST

Intransitive verbs are those which can appear with no complements at all. The verb *eat* has no complements in example 7.

7. He ate.

Transitive verbs take a single object complement. This complement may be a noun (direct object), a prepositional phrase, a finite complement, etc. *Eat* and *find* are transitive in 8a. and 8b. respectively. Ditransitive verbs have more than one object complement. *Give* and *lower* are ditransitive in 9a. and 9b. respectively. The verb-phrase complement of linking verbs is one that reidentifies the subject of the verb. *Become* is linking in 10. In complex-transitive verbs there are two verb-phrase complements, which may be in a predication relationship, shown in 11a. and 11b., or an identity relationship, found in 11c.

8a. He ate the cake.
8b. He found that I had eaten.

9a. John gave Mary the book.
9b. John lowered the price to $5.00

10. John became king.

11a. We painted the house purple.
11b. I wanted him to leave.

11c. They elected him president.

Verbs can, and often do, fall into more than one complementation class. For example, consider the verb *treat*.

12.
```
{base=treat
    entry=E0061964
        cat=verb
        variants=reg
        intran
        tran=np
        tran=pphr(with,np)
        tran=pphr(of,np)
        ditran=np,pphr(to,np)
        ditran=np,pphr(with,np)
        ditran=np,pphr(for,np)
        cplxtran=np,advbl
        nominalization=treatment|noun|E0061968
    }
```

See Section 5.1 on page 34 for details of verb complement coding.

Verb entries also encode each of the inflected forms, (principal parts of the verb) in a **variants=** slot. Verbs are inflectionally classified as regular, Greco-Latin regular or irregular. See "Verb Inflection" on page 11, for more detail.

1.4 Nouns

As described above, noun entries describe the inflection of the nouns (pluralization) in a **variants=** slot, and spelling variation in a **spelling_variant=** slot. The **compl=** slot indicates complementation for nouns. A **nominalization=** slot indicates that the noun is the nominalization of a verb or adjective. Nouns also have a **trademark=** slot to list tradenames, as well as a **trademark** code (with no argument) to indicate that a term is a trademark without an identifiable name.

1.5 Adjectives

In addition to inflection (**variants=**) codes and complement codes, adjectives in SPECIALIST have position codes, in a **position=** slot, to indicate the syntactic positions in which they occur.

Adjectives that occur pre-nominally in noun phrases are marked **attrib(),** in the **position=** slot. The numerical argument of the **attrib()** slot indicates where in the normal sequence of noun premodifiers this adjective occurs. Qualitative adjectives (**attrib(1)**) normally precede color (**attrib(2)**) and classifying (**attrib(3)**) adjectives. 13a. is more natural than either 13b. or 13c.

13a. a big red wooden box.
13b. a red big wooden box.

Adjectives that can occur in predicate adjective constructions have the code **pred** in their **position=** slot, and adjectives which can occur post nominally have the code **post**. See "Adjective Positions" on page 61.

1.6 Adverbs

Adverbs in SPECIALIST are coded to indicate their modification properties in a **modification_type=** slot. SPECIALIST recognizes sentence, verb-phrase and intensifier type adverbs, as well as classifying verb-phrase and sentence adverbs into manner, temporal and locative types. Adverbial particles like *up* in 14. are also listed as adverbs in SPECIALIST, with a modification_type indicating that it is a particle.

14. I called them up.

2. Spelling Variation

While spelling is highly standardized in Modern English, spelling variation remains fairly common. Some spelling variation is due to dialect differences, such as the well-known differences between British and American spelling conventions, especially in technical vocabulary. Table 2 describes some American and British English spelling differences.

American Spelling	British Spelling	Example
e	ae	hemo/haemo
e	oe	fetus/foetus
er	re	center/centre
ection	exion	inflection/inflexion
z	s	analyze/analyse

Table 2. American and British Spelling Differences

Many words show spelling variation in American English. For example*, artifact* has the spelling variant *artefact* listed in several modern American dictionaries (See Emery (1973)). Spelling variants when known are collected as the fillers of the **spelling_variant=** slots in lexical records.

3. Syntactic Category (Part Of Speech)

Each entry includes a **cat=** slot, showing the syntactic category of the entry. Table 3 shows the allowable fillers of the **cat=** slot, the syntactic categories they represent and some examples of each category.

Code	Category	Examples
cat=verb	Verbs	*see, run, anaesthetize*
cat=aux	Auxiliary Verbs	*do, have, be*
cat=modal	Modal Auxiliaries	*may, can, shall, could*
cat=noun	Nouns	*boy, milk, surgery*
cat=pron	Pronouns	*he, she, it, they*
cat=adj	Adjectives	*red, optical*
cat=adv	Adverbs	*quickly, fast, probably, up*
cat=prep	Prepositions	*in, of, on, in regard to*
cat=conj	Conjunctions	*and, or, but*
cat=compl	Complementizers	*that*
cat=det	Determiners	*this, that, these, those*

Table 3. SPECIALIST Syntactic Categories

4. Variants: Agreement and Inflection

The **variants=** slot records inflectional and agreement information. Each entry has at least one **variants=** slot indicating inflectional morphology and/or agreement facts about the entry.

4.1 Verb Inflection

The **variants=** slot records the verb's inflectional pattern. English main verbs have five forms (principal parts): the base form (infinitive), third person singular form, past tense form, present participle form and past participle form. When a verb adheres to the regular English inflection pattern, it is marked **reg** or **regd**. The code **reg** indicates regular inflection as defined in "Regular Verb Inflection" on page 12; **regd** indicates that the final consonant is doubled. (See "Regular Doubling Inflection" on page 13.) If the verb is irregular it receives the code **irreg**| | | | |. The actual inflections are recorded in the **irreg** code, See Section 4.1.3 on page 14.

SPECIALIST inflection codes refer to the *spelling*s of lexical items, not to their phonology. A lexical item which is phonologically regular may be orthographically irregular. Although it is extremely rare for an English verb to have a phonologically irregular present participle form, there are verbs whose present participle is orthographically irregular. For instance, the present participle of *glue* can be spelled regularly (*gluing*) or irregularly (*glueing*) with the same regular phonology.

4.1.1 Regular Verb Inflection

The filler **reg** is added to the **variants**= slot of regular verbs. Verbs are considered regular if they meet the following description:

1. The third person present tense singular suffix is *s.*
   y becomes *ie* following a consonant before the suffix *s*.
   e is inserted between a base ending in *z*, *x*, *ch*, or *sh* and the suffix *s*.
2. The past tense and the past participle suffix is *ed*.
   y becomes *ie* following a consonant before the suffix *ed.* final
   e is deleted before the suffix *ed*.
3. The present participle suffix is *ing*.
   ie becomes *y* before the suffix *ing.*
   final e is deleted before the suffix *ing*,
   unless preceded by *e*, *y*, or *o*.

The alternation of *y* with *ie* and the dropping of the silent *e* interact in forming the past tense/past participle of regular verbs. The {*y* ~ *ie*} alternation precedes *e*-dropping. For example, in the past tense/past participle of the verb *fry*, *y* becomes *ie* and the final *e* is dropped to produce *fried*, rather than **frieed*. (The asterisk preceding **frieed* indicates an ungrammatical or ill-formed utterance.) And *ie* becomes *y* before the final *e* is deleted in the present participle of *tie*, producing *tying* rather than **tiing.* Notice, too, that the final *e*-dropping rule applies differently to the past tense/past participle and the present participle. The final *e* of *hoe*, for example, is dropped before *ed* and retained before *ing*, i.e. *hoeing* and *hoed*. Table 4 illustrates the regular pattern of verb inflection.

Base Ends with:	**3rd Singular Ends with:**	**Past / Past Participle ends with:**	**Present Participle ends with:**	**Example paradigms:**
-s	-ses	-sed	-sing	dismiss: dismisses, dismissed, dismissing
-z	-zes	-zed	-zing	waltz: waltzes, waltzed, waltzing
-x	-xes	-xed	-xing	index: indexes, indexed, indexing
-ch	-ches	-ched	-ching	detach: detaches, detached, detaching
-sh	-shes	-shed	-shing	distinguish: distinguishes, distinguished, distinguishing
-ie	-ies	-ied	-ying	tie: ties, tied, tying
-ee	-ees	-eed	-eeing	agree: agrees, agreed, agreeing
-oe	-oes	-oed	-oeing	canoe: canoes, canoed, canoeing
-ye	-yes	-yed	-yeing	dye: dyes, dyed, dyeing
-*C*y[a]	-*C*ies	-*C*ied	-*C*ying	dry: dries, dried, drying

Table 4. Regular Verb Inflection

Base Ends with:	3rd Singular Ends with:	Past / Past Participle ends with:	Present Participle ends with:	Example paradigms:
-*D*e[b]	-*D*es	-*D*ed	-*D*ing	love: loves, loved, loving
-*X*[c]	-*X*s	-*X*ed	-*X*ing	talk: talks, talked, talking

Table 4. Regular Verb Inflection

a. *C* denotes any consonant or consonant cluster.
b. *D* denotes any letter other than *i*, *y*, *e*, or *o*.
c. *X* denotes any final letter not otherwise covered in the table.

4.1.2 Regular Doubling Inflection

The filler **regd** is added to the **variants=** slot of verbs displaying regular doubling inflection. Cummings (1988) gives the following rule for orthographic consonant doubling in English:

> The final consonant letter of a stem is twinned only when all of the following conditions are met:
>
> 1. The stem must be free.
> 2. The stem must end in VC# -- that is, with a single vowel letter that is followed by a single consonant letter that spells a single consonant sound.
> 3. The VC# string in the stem must bear primary or secondary stress both before and after the suffix is added.
> 4. The suffix being added starts with a vowel.
> 5. The suffix must not be one of the shortening suffixes, such as *-ic* or *-ity.*

These rules, especially rule three and the "spells a single consonant" clause of rule 2, refer to the phonology of the verb. Since the SPECIALIST system has no access to phonology, a purely orthographic approximation of these rules is used in the SPECIALIST lexicon coding scheme.

Verbs ending in an orthographic CVC pattern, whose final consonant is doubled before the past tense and participle suffix *-ed* and the present participle suffix *-ing* but are otherwise regular are given the code **regd**, e.g. *bat*: *bats*, *batted*, *batting*.

The requirement that the verb end in a closed syllable with a single (orthographic) vowel means that verbs like *acquit*, *dial*, *duel*, *equip*, *fuel*, and *quit* are considered irregular.

Since consonant doubling generally occurs following a stressed vowel, many of these verbs are one syllable. But multi-syllable verbs with final consonant doubling do exist, e.g. *commit*, *control*, *overlap*, *transfer* and *debug*, even when the vowel of the final syllable is unstressed

as in *level* and *bootleg*; *level*: *levels*, *levelled*, *levelling*, and *bootleg*: *bootlegs*, *bootlegged*, *bootlegging*. These are all considered regular doubling.

We do not consider addition of *k* following a final *c* to be an instance of regular doubling, although some scholars identify those processes. So, verb entries for *mimic* and *traffic* are irregular: *traffic*: *traffic*, *traffics*, *trafficked*, *trafficking*, and *mimic*: *mimic*, *mimics*, *mimicked*, *mimicking*.

Regular doubling applies to only the past tenses, past participles, and present participles of verbs; we do not consider the regular doubling rule to apply in the rare cases where a final *s* or *z* doubles before the third person singular present tense suffix *-es*. The verb *bias*, for example, is both regular and irregular but not regular doubling. Its two paradigms are: *bias*: *biases*, *biased*, *biasing* and *bias*: *bias*, *biasses*, *biassed*, *biassing* which are regular and irregular respectively. The verb *bus* is both regular doubling and irregular because its third person singular present tense form can be either *buses* or *busses*; irregular: *bus*: *bus*, *busses*, *bussed*, *bussing*; regular doubling: *bus*: *buses*, *bussed*, *bussing*.

Consonant doubling is often subject to dialectal or simple spelling variation; *travel* can be either regular or regular doubling in American English but it is regular in British English. Verbs like those are marked both regular and regular doubling.

Some cases in which the base form shows dialectal or other spelling variation involving doubling do not count as instances of regular doubling. The verb *program* has a British English spelling variant *programm*e. The result is that this verb has the following paradigm: *programs*/*programmes*; *programmed*/*programed*; *programming*/*programing*. While *programming* might appear to be the result of regular doubling applied to *program*, it is identical to the regular present participle of *programme* so the lexical record is coded as regular.

4.1.3 Irregular Verb Inflection

The **variants=** slot of irregular verbs is filled with the code **irreg| | | | | |** , with the irregular inflectional forms listed between the pipe "|" symbols in this order: base form, third person present tense, past tense, past participle followed by present participle. The filler of the **variants=** slot for the irregular verb *break* is: irreg|break|breaks|broke|broken|breaking|. Many of the verbs listed in the SPECIALIST lexicon as irregular are members of the class of English strong verbs; verbs with inflectional vowel changes and past participles which differ from their past tenses. e.g. *eat*: *eats*, *ate*, *eaten*, *eating*. Some verbs are nearly regular but fail to meet the rules given for regular or regular doubling above. e.g. *singe: singeing* which does not drop *e* before *ing*. And *stymie* does not undergo $\{ie \sim y\}$-alternation; *stymie ~ stymieing*. Such verbs are listed as irregular. Verbs ending in *o* which take *es* in the present tense, like *veto,* are considered irregular.

Verbs whose spelling variants differ in their inflectional paradigms are coded as irregular. The verb *fulfil* has the variant spelling *fulfill*. Since all variant codes, except **irreg| | | | | |**, apply to all spelling variants of a lexical record, *fulfil*/*fulfill* must be listed with two irregular codes, despite the fact that the two paradigms are, individually, regular and regular doubling; *fulfil*: *fulfil*,*fulfils*, *fulfilled*,*fulfilling* and *fulfill: fulfill*,*fulfills*,*fulfilled*,*fulfilling*.

Verbs with defective paradigms are also coded as irregular. The verb *sight-see* has a base form, and a present participle, but the other principal parts do not exist, **sight-sees, *sight-saw*, **sight-seen*. Similarly *beware* occurs only in its base form. The missing parts of the paradigm of these verbs are indicated by leaving the position for them in the **irreg**| | | | | code empty.

4.2 Variants of Modals and Auxiliaries

Modal and auxiliary verbs differ from main verbs in the richness of their inflectional paradigm. *Be* has more inflections than most verbs and the modals have fewer. Modals and auxiliaries also have cliticized and negative contracted forms. This variation is captured in the **variant=** slot. The **variants=** slot found in the entries for most verbs can be thought of as an abbreviation of several **variant=** slots.

The fillers of the **variant=** are the variants themselves with features attached following a semi-colon. The main part of a variant feature is a tense code, indicating the tense (past or present) of the variant. The tense codes take arguments indicating agreement restrictions on the variant; no argument means that agreement is unrestricted. The agreement features are the same ones used to describe pronoun agreement.

Tense Code	(	List of Agreement Features	)	:	Negation Code
past pres		free fst_plur second sec_sing sec_plur third thr_sing thr_plur			negative
infinitive past_part pres_part pres past					

Table 5. Features in the variant= slot

4.2.1 Tense in Modals and Auxiliaries

The modal verbs *can*, *may*, *shall* and *will* have past tense forms *could*, *might*, *shall*, and *would*. While these forms are not semantically identical to past tense in main verbs, they function syntactically as past tense in some cases. For example, 15a. could be a report that the subject of 15a. uttered 15b. as well as 15c. This back-shifting phenomenon in indirect quotations is most

naturally stated in terms of past tense, since the difference between *hope* and *hoped* in 16b. and 16c. is uncontroversially a matter of tense.

15a. He said he would go.
15b. I will go.
15c. I would go.

16a. He said he hoped to attend the meeting.
16b. I hope to attend the meeting.
16c. I hoped to attend the meeting.

Past and present tense modals are grouped together in the same entry with the past and present forms appearing in separate **variant**= slots. The features **past** and **pres** indicate the past and present form of the modal respectively. These codes also allow SPECIALIST to capture the special agreement behavior of the modal verbs. *May* agrees with any noun or pronoun subject regardless of person or number as **variant=may;pres** indicates. Similarly, *might* agrees with any noun or pronoun as **variant=might;past** indicates.

17.
```
{base=may
    entry=E0039142
        cat=modal
        variant=might;past
        variant=mayn't;pres:negative
        variant=mightn't;past:negative
    }
```

The single variant of *must* is considered to be present tense.

18.
```
{base=must
    entry=E0041474
        cat=modal
        variant=mustn't;pres:negative
    }
```

4.2.2 The Paradigm of *be*

The auxiliary verb *be* has a richer inflectional paradigm than other verbs. Unlike *have* and *do,* these forms could not be fitted into an **irreg| | | | |** filler of a **variants=** slot. *Be* has no form equivalent to the past tense of a main verb. *Ate*, for example, agrees with any subject without regard to person or number, but *was* agrees only with first and third person singular subjects and *were* agrees with second and third person plural subjects as well as first person plural subjects.

19a. *We/they was going home.
19b. We/they ate at three o'clock.
19c. We/they were going home
19d. *I/he were going home.

Each of the inflectional forms of *be* (*am*, *is*, *was*, *are*, *were*, *been*, *being*) is given in a **variant**= slot with the appropriate agreement features. Cliticized and negative forms are also given. See the discussion in 4.2.4 and 4.2.5 below.

20.
```
{base=be
    entry=E0012152
```

```
        cat=aux
        variant=be;infinitive
        variant=is;pres(thr_sing)       variant='s;pres(thr_sing)
        variant=isn't;pres(thr_sing):                negative
        variant=are;pres(fst_plur,second,thr_plur)
        variant='re;pres(second,thr_plur)
        variant=aren't;pres(fst_plur,second,thr_plur):negative
        variant=am;pres(fst_sing)
        variant='m;pres(fst_sing)
        variant=was;past(fst_sing,                   thr_sing)
        variant=wasn't;past(fst_sing,          thr_sing):negative
        variant=were;past(fst_plur,second,thr_plur)
        variant=weren't;past(fst_plur,second,thr_plur):negative
        variant=been;past_part
        variant=being;pres_part
    }
```

4.2.3 The Paradigms of *do* and *have*

The inflectional variants of the auxilliary verbs *do* and *have* are listed in **variant=** slots. The main verbs *do* and *have* are treated as ordinary irregular verbs.

21.
```
{base=do
    entry=E0023651
        cat=aux
        variant=do;pres(fst_sing,fst_plur,second,thr_plur)
        variant=don't;pres(fst_sing,fst_plur,second,thr_plur):negative
        variant=does;pres(thr_sing)
        variant=doesn't;pres(thr_sing):negative
        variant=did;past
        variant=didn't;past:negative
    }
```

22.
```
{base=have
    entry=E003089
        cat=aux
        variant=have;infinitive
        variant=have;pres(fst_sing,fst_plur,second,thr_plur)
        variant=has;pres(third)
        variant=had;past
        variant=having;pres_part
        variant=hadn't;past:negative
        variant=hasn't;pres:negative
        variant=haven't;pres(fst_sing,fst_plur,second,thr_plur):negative
        variant='ve;pres(fst_sing,fst_plur,second,thr_plur)
        variant='d;past
    }
```

4.2.4 Clitic Forms

In written English tensed auxiliary and modal verbs can be contracted onto the subject noun phrase.

23a. He's going to the picnic.
23b. He'll be leaving soon.
23c. He'd like to take us along.

These cliticized forms are recorded in the **variant=** slot of auxiliary and modal verbs. For example, in the modal entry for *will* in example 24., the clitic forms *'ll* and *'d* are recorded in **variant=** slots, with the same agreement features as *will* and *would* respectively.

24.
```
{base=will
    entry=E0065450
        cat=modal
        variant=would;past
        variant=won't;pres:negative
        variant=wouldn't;past:negative
        variant='ll;pres
        variant='d;past
    }
```

4.2.5 Negative Contracted Forms

In written English *not* may be contracted onto the tensed auxiliary (including modal auxiliaries).

25a. He isn't coming to the picnic.
25b. He won't go with us.
25c. He didn't even try.

These contracted forms might be thought of as involving a variant form of the adverb *not*. But, since the form of the contraction depends on the auxiliary, it is best to regard these as negative variants of the auxiliaries. *Can* has both *can't* and *cannot*, while *will* has *won't*, not **willn't.* In SPECIALIST these contracted forms are recorded in **variant=** slots. Since inflected forms of the auxiliary are contracted, there may be several **variant=** slots containing negative contractions in an entry; *can*, for example, has **variant=can't, variant=cannot**, and **variant=couldn't** as its lexical entry shows.

26.
```
{base=can
    entry=E0014877
        cat=modal
        variant=could;past
        variant=couldn't;past:negative
        variant=cannot;pres:negative
        variant=can't;pres:negative
}
```

The feature **negative** in the negative contracted forms represents strict negation, as it does in the entries for strictly negative adverbs. See "True Negative Adverbs" on page 71.

4.3 Adjective Inflection (Comparison)

4.3.1 Regular Variants

Regular adjectives form their comparative and superlative according to the following rules: The comparative suffix is *er*.

Following a consonant, *y* becomes *ie* before the *er* suffix.

Final *e* is deleted before the *er* suffix.

The superlative suffix is *est*.

Following a consonant before the *est* suffix, *y* becomes *ie*.

Final *e* is deleted before the *est* suffix.

As with verbs {*y* ~ *e*} alternation precedes final *e*-drop, so in words like *dizzy*, *y* becomes *ie* and the final *e* is subsequently dropped; it's *dizzier* not **dizzieer*.

These rules give the pattern shown in Table 6.

Base ends with:	Comparative ends with:	Superlative ends with:	Example Paradigms
-C[a]y	-Cier	-iest	brainy: brainier, brainiest
-V[b]y	-Vyer	-Vyest	gray: grayer, grayest
-Ce	-Cer	-Cest	fine: finer, finest
-Ve	-Ver	-Vest	blue: bluer, bluest
-X[c]	-Xer	-Xest	clear: clearer, clearest

Table 6. Adjective Inflection

a. C stands for any Consonant letter.
b. V stands for any Vowel letter.
c. X stands for any letter other than y or e.

Regular Adjectives have the code **reg** in their **variants=** slot.

4.3.2 Regular Doubling

Regular doubling adjectives follow the regular pattern except that they double the final consonant before the suffixes *er* and *est,* e.g. *fat*, *fatter*, *fattest.*

Regular doubling adjectives have the code **regd** in their **variants=** slot.

4.3.3 Irregular Adjectives

Any adjective which inflects in a way which does not conform to the rules for regular or regular doubling adjectives is considered irregular. The superlative and comparative of irregular forms are listed inside the code **irreg| | | |**; the base form first, followed by the comparative and the superlative. Irregular adjectives include superlative cases like: *good*: *better*, *best* as well as cases which deviate more subtly from the rules for regular and regular doubling adjectives. E.g. *shy*, *spry* and *sly* violate *{y ~ ie}* alternation: *shy*: *shyer*, *shyest*. In *gooey* and *pricey*, *ey* alternates with *ie*: *gooey*: *gooier*, *gooiest*. *Old* has two paradigms, one regular and one irregular: *old*: *older*, *oldest*, and *old*: *elder*, *eldest*. *Far* has two irregular paradigms: *far*: *farther*, *farthest* and *far*: *further, furthest*.

4.3.4 Invariant Adjectives

Invariant adjectives have no morphological comparative or superlative form. This includes non-gradable adjectives like *medical* or *daily* as well as periphrastic adjectives which compare with *more* or *most*. **inv** in the **variants=** slot of an adjective indicates that the adjective is invariant. Note that **inv** for adjectives and adverbs has a meaning quite different from **inv** for nouns. See "Invariant Nouns" on page 22.

27.
```
{base=acoustic
    entry=E0006949
        cat=adj
        variants=inv
        position=attrib(3)
        position=pred
        stative
    }
```

4.3.5 Periphrastic Adjectives

Periphrastic adjectives form their comparative and superlative with *more* and *most*. Since in standard English adjectives in periphrastic constructions do not inflect morphologically (**more beautifuller*), periphrastic adjectives are necessarily invariant. The code for periphrastic adjectives is **periph**. When an invariant adjective (**inv**) is periphrastic, it receives one **variants=** slot of the form **variants=inv; periph.**

4.4 Adverb Inflection (Comparison)

The variants of adverbs are similar to those of adjectives.

4.4.1 Regular Adverbs

Adverbs that form their comparative and superlative according to the regular pattern of English adverb inflection are considered regular, and their **variants=** slot contains the code **reg**.

Regular adverbs form their comparative and superlative according to the following rules: The

comparative suffix is *er*.

Following a consonant, *y* becomes *ie* before the *er* suffix.

Final *e* is deleted before the *er* suffix.

The superlative suffix is *est*.

Following a consonant before the *est* suffix, *y* becomes *ie*.

Final *e* is deleted before the *est* suffix.

{*y* ~ *ie*} alternation precedes final *e*-drop.

These rules are the same as the rules for regular adjective inflection: *early* (positive), *earlier* (comparative), *earliest* (superlative).

28.
```
{base=early
    entry=E0024315
        cat=adv
```

variants=reg
modification_type=verb_modifier;temporal
}

4.4.2 Irregular Adverbs

When an adverb has an irregular comparative and/or superlative form, an **irreg| | | |** filler is used in the **variants=** slot. The **irreg| | | |** has three argument positions, the first for the base form, second for the comparative and the third for the superlative form. *Well* is an irregular adverb: *well* (positive), *better* (comparative), *best* (superlative).

29.
```
{base=well
    entry=E0065231
        cat=adv
        variants=irreg|well|better|best|
        modification_type=verb_modifier;manner
    }
```

4.4.3 Invariant Adverbs

Adverbs that have no comparative or superlative form have **inv** in their **variants=** slots. The code **inv** has a different meaning in adjective and adverb entries than it does in noun entries. See "Invariant Nouns" on page 22.

30.
```
{base=always
    entry=E0008403
        cat=adv
        variants=inv
        modification_type=intensifier
        modification_type=verb_modifier;temporal
    }
```

4.4.4 Periphrastic Adverbs

Adverbs which form the comparative and superlative with *more* and *most* have the code **periph** following a semi-colon after the **variants=inv** code. *Often* is a periphrastic adverb: *often* (positive), *more often* (comparative), *most often* (superlative).

31.
```
{base=often
    entry=E0043653
        cat=adv
        variants=inv;periph
        modification_type=verb_modifier;temporal
    }
```

4.5 Noun Inflection

4.5.1 Countability

A major distinction is made in the SPECIALIST lexicon between count and uncount nouns, corresponding to the traditional categories of countable, abstract and mass nouns. Both abstract and mass nouns are considered uncount. Generally, nouns are considered count if they have distinct singular and plural forms which agree with singular and plural verbs respectively. Some nouns are invariant in form, but may be count (e.g., *sheep*, shown in 33a-d.). Count nouns can be determined by numbers, *a*/*an*, *many*, etc., and they cannot occur in the singular with a zero determiner, whereas uncount nouns are the opposite (as in 34.)

32a. A book is on the desk.
32b. Two books/many books are on the desk. 32c.
*Book is on the desk. Books are on the desk.

33a. A sheep is in the field.
33b. Two/many sheep are in the field.
33c. Sheep is in the field.
33d. Sheep are in the field.

34a. * A sand is on the beach.
34b. *Two /many sands are on the beach.
34c. Sand is on the beach.
34d. *Sand are on the beach.

Uncount nouns are represented by the fillers **uncount** and **groupuncount** in the **variants=** slot; count nouns are indicated by **reg**, **glreg, metareg**, **irreg**, **sing**, **plur**, **inv**, and **group()**. Each of those codes is discussed below.

4.5.2 Regular Nouns

When a noun follows the regular pattern of English plural formation, the **variants=** slot contains the filler **reg**.

Nouns are considered regular if they conform to the following rules:

1. The plural suffix is *s*.
2. *y* becomes *ie* following a consonant before the *s*.
3. *e* is inserted before the plural suffix *s* if the base ends in *s*, *z*, *x*, *ch*, or *sh*.

The results of these rules can be summarized in Table 7.

Base ends with:	Plural ends with:	Examples
-Cy [a]	-Cies	fly: flies
-s	-ses	illness: illnesses
-z	-zes	waltz: waltzes
- x	-xes	box: boxes
-ch	-ches	match: matches
-sh	-shes	splash: splashes
-X [b]	-Xs	book: books

Table 7. Regular Noun Inflection

a. C stands for any consonant letter.
b. X stands for any termination other than *y, s, z, x, ch* or *sh*.

No other nouns are considered regular in this system.

Since these are orthographic rules which incompletely reflect phonological rules, some words which might be considered regular in English are irregular in this system. The word *stomachs*, for example, is considered an irregular plural in this system, since rule 3. does not distinguish "hard" from "soft" *ch*'s.

The rules above consider only the end of a term, whether or not the term consists of several words. So left headed multi-word terms like *body politic*, *court martial* and *notary public* which form their plural according to regular English rules applied to the first word of the term (*bodies politic*, *courts martial* and *notaries public),* are considered irregular. Similarly, *passersby*, *tablespoonsful*, and *filets mignons* are irregular.

Although many English words ending in *o* take *es* in the plural, the rules above categorize them as irregular, e.g *buffaloes*, *potatoes*, *vetoes*, and *volcanoes*. Similarly, words ending in *uy* often participate in *{y ~ ie}* alternation, but since rule 2 requires a preceding consonant these words will be considered irregular (e.g. *colloquies*, *soliloquies*, *obsequies*).

Those few nouns which double the final consonant before the plural affix, (*busses*, *quizzes*, *fezzes*), are considered irregular; there is no **regd** code for nouns.

4.5.3 Greco-Latin regular plurals

Words of Classical origin, which are common in the biomedical domain, often retain their Latin or Greek inflectional pattern in English. SPECIALIST therefore encodes these words with the **glreg** filler of the **variants=** slot to indicate Greco-Latin inflection.

Nouns are considered Greco-Latin regular if they follow one of the paradigms illustrated in Table 8.

singular ends with:	plural ends with:	Examples
-us	-i	focus/foci
-ma	-mata	trauma/traumata
-a	-ae	larva/larvae
-um	-a	ilium/ilia
-on	-a	taxon/taxa
-sis	-ses	analysis/analyses
-is	-ides	cystis/cystides
-men	-mina	foramen/foramina
-ex	-ices	index/indices
-x	-ces	matrix/matrices

Table 8. Greco-Latin Plural Nouns

These rules cover only a few common Greco-Latin patterns of noun inflection. Words like *corpus*

(pl. *corpora*) are not considered Greco-Latin regular in our system. As with the regular inflection rule these rules apply only to the end of a term. Multi-word terms consisting of a Latin noun followed by a Latin adjective are not Greco-Latin plural even if both the noun and adjective inflect according to the Greco-Latin paradigm given in Table 8. For example, the plural of *Lactobacillus fermentum* is *Lacotbacilli fermenta*.

In cases of apparent conflict between rules, *ma ~ mata* vs. *a ~ ae, sis ~ ses* vs. *is ~ ides* and *ex ~ ices* vs. *x ~ ces*, the one with the longer singular ending is used. That is, the plural of any word marked **glreg** ending in *ma* ends in *mata* not *mae* and the plural of a glreg word ending in *sis* ends in *ses* not *sides*. The rare cases in which a word ending in *ma* does have a plural ending in *mae*, are treated as irregular. For example, *lacrima*, *mamma* and *squama* are irregular, not Greco- Latin regular in our system. Their plurals are *lacrimae*, *mammae* and *squamae*. Similarly, *axis* (plural *axes*) is irregular.

4.5.4 Meta-linguistic regular nouns

The plural of acronyms, in which the constituent letters of the orthography are pronounced as their letter name, numbers and other orthographically meta-linguistic nouns may form their plural with an apostrophe s. There is considerable individual variation in this, so the plural of *Ph.D.* may be written *Ph.D.'s* or *Ph.D.s*. Similarly the plural of *5* could be *5's* or *5s*. The **variants=** slot filler **metareg** is used in SPECIALIST to capture this variation. Nouns (most often acronyms) coded as **metareg** can have a plural with simple *s* or with *'s*.

4.5.5 Irregular Nouns

The plural form for irregular nouns is explicitly listed in an **irreg| | |** filler of the **variants=** slot. The irregular plural form is given as the second argument of the **irreg| | |** code, as illustrated in the lexical item for *calf* below.

34.
```
{base=calf
    entry=E0014750
    cat=noun
        variants=irreg|calf|calves|
        variants=uncount
        compl=pphr(of,np|leg|)
    }
```

4.5.6 Fixed Singular Nouns

Some English nouns behave like count nouns but lack a plural form. These fixed singular forms are indicated by the filler **sing** in the **variants=** slot.

Nouns are considered fixed singular if they meet the following criteria:

1) They lack a plural form.
2) They agree with a singular verb.
3) They may be determined by a/an but do not appear with numerical determiners.
4) They do not appear with zero determiner.

These traits are illustrated by the fixed singular noun *lope*, 'an easy swinging gate'.

35a. He jogged through the room at a lope.
35b. *They jogged through the room at two lopes. 35c.

*Lope is a poor way to get home.

4.5.7 Fixed Plural Nouns

Fixed plural nouns, like fixed singular nouns, are basically count nouns. Fixed plurals are count nouns that lack a singular form. They are indicated in the **variants**= slot by the filler **plur.**

Nouns are considered fixed plural if they:

1) Agree only with plural verbs, and
2) have no distinct singular form.

The following examples illustrate those properties.

36a. The cattle are grazing.
36b. *A cattle is grazing.
36c. We saw those cattle.
36d. *We saw this cattle.

37a. We called the police.
37b. *We called a police.

38a. The surroundings were beautiful.
38b. *The surrounding was beautiful.
38c. These surroundings are beautiful.
38d. *This surrounding is beautiful.

Many fixed plurals seem to show regular plural morphology (e.g. *auspices*, *amends*, *entrails*, *guts*, *odds*, *remains*, *regards*, *particulars*, *premises*, etc.) but they are considered fixed plural rather than regular, since they lack a singular form.

40.
```
{base=surroundings
    entry=E0059344
        cat=noun
        variants=plur
    }
```

4.5.8 Invariant Nouns

Nouns are considered invariant if they have the same form in the singular and the plural (e.g. *sheep*, *means*), but remain countable. The **variants=** slot for these nouns is filled with the code **inv.** These nouns should not be confused with fixed singular, fixed plural, or uncount nouns which have only one form. Invariant nouns may be thought of as having both a singular and a plural form that happen to be the same.

41a. A sheep has been found.
41b. Five sheep have been found.

42a. The quickest means of travel is by plane.
42b. The means justify the ends.

43. There are five deer by the pond.

44.
```
{base=deer
    entry=E0021150
        cat=noun
```

variants=inv
}

Words whose singular and plural are pronounced differently but spelled the same are coded invariant, e.g. *corps*, *Sioux*, *chamois*, etc.

Invariant nouns sometimes have homophonous (homographous) regular nouns, meaning "type of" the invariant noun.

45a. There are three fish in this bowl.
45b. There are four carnivorous fishes in this region.

Some terms denoting animals or fish are either invariant or regular depending on whether they are considered "game".

46a. The hunters bagged three bear.
46b. We saw three bears in the woods.
46c. The fishermen caught six carp.
46d. We saw three carps in the pond.

In such cases there are two **variants=** slots, one invariant and the other regular.

Invariant nouns are not limited to the common English vocabulary or to the sub-language of hunters. Table 9 lists some invariant nouns in the biomedical domain.

Singular	**Plural**
aditus	aditus
aquaeductus	aquaeductus
arcus	arcus
decubitus	decubitus

Table 9. Some invariant nouns as listed in Dorland's Illustrated Medical Dictionary.

Singular	**Plural**
descensus	descensus
ductus	ductus
exitus	exitus
facies	facies
ictus	ictus
introitus	introitus
lacus	lacus
manus	manus
nexus	nexus
processus	processus
pulsus	pulsus
recessus	recessus
situs	situs
tractus	tractus

Table 9. Some invariant nouns as listed in Dorland's Illustrated Medical Dictionary.

4.5.9 Group Nouns

Group nouns have a **group()** filler in their **variants=** slot. The argument of the **group()** code indicates the inflectional morphology of the noun, while **group()** refers to the agreement facts of the noun. The allowable arguments of the **group()** code are: **reg**, **irreg**, **sing**, **glreg**, and **metareg**.

The term **group** corresponds, generally, to the traditional term *collective*. The singular form of group nouns is indeterminate as to number; it can agree with either a singular or plural verb. Collective nouns are much more common in British English than American English.

47a. The committee has met.
47b. The committee have met.

48a. My family are all Socialists.
48b. My family always votes Socialist.

49a. The majority of the class are freshmen. 49b.
The majority of the class takes Spanish.

The meaning difference between these pairs of examples is said to be "point of view". See Quirk et al. (1985, section 5.108), for further discussion and examples.

Some group nouns take a prepositional (*of*) phrase complement which indicates what the group denoted by the group noun consists of.

50a. A herd of cattle is/are munching grass in the meadow. 50b.
The board of directors is/are angry over the attempted take over.
50c. A bunch of flowers was/were lying on the floor.
50d. A tall stack of pancakes was/were his favorite breakfast.

Some group nouns are uncount, or fixed singular. They have no separate plural form, and the form they have is indeterminate between singular and plural. These nouns have the code **groupuncount** in their **variants=** slot.

51a. The new right are a bunch of dangerous hooligans.
51b. The new right is gaining votes in the south.

See "Group Uncount Nouns" below.

4.5.10 Uncount Nouns

Uncount nouns are indicated by the code **uncount** as filler for the **variants=** slot.

The term uncount corresponds to the traditional terms abstract and mass. These nouns have no separate plural form, and unlike group (count) or invariant nouns agree only in the singular. Only uncount nouns may appear in the singular with a zero determiner.

The examples below illustrate these properties for the nouns *sincerity* and *dirt*: 52a.

Sincerity is hard to fake.
52b. *(Five) sincerities is/are hard to fake.
52c. Dirt is not good to eat.
52d. *(Five) dirts is/are not good to eat.

53. {base=dirt
entry=E0023123

```
        cat=noun
        variants=uncount
}
```

Some uncount nouns have homophonous (homographous) count nouns which mean "kinds, units, instances," etc. of the uncount noun.

54a. He drank a lot of beer. (uncount)
54b. He drank five beers. (count (= bottles of beer))

55a. Arthritis is a painful disease. (uncount)
55b. There are several different arthritides.
(count (= types of arthritis))

56a. Truth is the core of social morality. 56b.
We hold these truths to be self evident.

Lexical entries for nouns like this have **uncount** and **reg, irreg** or **glreg** in their **variants=** slot.

4.5.11 Group Uncount Nouns

Uncount nouns, and fixed singular nouns which are group, are given the code **groupuncount** in their **variants=** slot. These nouns have no distinct plural form, but the singular agrees as both singular and plural. They differ from invariant nouns in that they cannot be determined by *a, an* or a number.

Group uncount includes unique uncount terms like *United States*, *Kremlin* and *parliament*, generic terms like *intelligentsia*, *laity* and *faculty*, as well as collective terms like *mankind* which simply lack a plural.

57a. The parliament is/are meeting today.
57b. Mankind is/are the main environmental problem.
57c. The intelligentsia is/are skeptical of the new order.

We consider words ending in *ics* which are indeterminate with respect to agreement to be group uncount (e.g. *heuristics*, *optics*, *politics* and *statistics* but not *linguistics* or *physics*.)

58a. His politics is/are of no interest to me.
58b. Linguistics is not an exact science.
58c. *Linguistics are not an exact science.

4.6 Agreement for Pronouns

In English, pronouns show agreement (concord) in terms of both number and person.

The **variants=** slot for pronouns records the person and number features needed for agreement. Three persons (First, Second, and Third) and two numbers (Singular and Plural) are combined in the six fillers for the **variants=** slot for pronouns.

	Singular	Plural
First Person	fst_sing	fst_plur
Second Person	sec_sing	sec_plur
Third Person	thr_sing	thr_plur

Table 10. Person / Number Codes For Pronouns

Since there is so much syncretism of the singular and plural forms of second and third person pronouns, the codes **second** and **third** have been added to denote those pronouns which have syncretic singular and plural forms. The code **free** was also added to denote a pronoun which agrees with all person and number combinations. The result is summarized in Table 11.

Slot	Fillers
variants=	free fst_plur fst_sing sec_plur sec_sing second third thr_plur thr_sing

Table 11. Fillers of the **variants=** slot for pronouns.

These same codes are used as features in the **variant=** slot of auxiliary and modal verbs. See "Variants of Modals and Auxiliaries" on page 11.

4.7 Agreement for Determiners

Determiners are distinct from adjectives because they exhibit agreement with nouns in English. The **variants=** slot for determiners (**cat=det**) gives the number characteristics of the nouns that they determine; for example, *a* can only determine a singular count noun.

59a. *I need a mud. (uncount)
59b. *I need a boxes. (plural)
59c. I need a box. (singular)

And, *both* can only determine a plural (count) noun.

60a. *Both mud are mine. (uncount)
60b. *Both box are mine. (singular)
60c. Both boxes are mine. (plural)

There are six fillers for the **variants=** slot of determiners.

Slot	Filler
variants=	sing plur uncount singuncount pluruncount free

Table 12. Fillers of the variants slot for determiners

The fillers of the **variants=** slot for determiners are discussed in sections 4.7.1 through 4.7.6 below.

4.7.1 Determiners of Singular Nouns

Determiners with the code **sing** can only determine singular count nouns. The examples in 61a. through 61c. show that *each* meets this criterion.

61a. Each boy did well.
(count singular)

61b. *Each dirt was on the floor.
(uncount)

61c. *Each boys did well.
(count plural)

62.
```
{base=each
    entry=E0024304
        cat=det
        variants=sing
    }
```

4.7.2 Determiners of Plural Nouns

Determiners with the code **plur** can only determine plural count nouns. The examples in 63a. through 63c. show that *many* meets this criterion.

63a. *Many boy did well.
(count singular)

63b. *Many dirt was on the floor.
(uncount)

63c. Many boys did well. (count
plural)

64.
```
{base=many
    entry=E0038864
        cat=det
        variants=plur
    }
```

4.7.3 Determiners of Uncount Nouns

Determiners with the code **uncount** can only determine uncount (aka mass) nouns. The examples in 65a. through 65c. show that *much* meets this criterion.

65a. *Much boy did well.
(count singular)
65b. Much dirt was on the floor.
(uncount)
65c. *Much boys did well.
(count plural)

66.
```
{base=much
    entry=E0041165
        cat=det
        variants=uncount
    }
```

4.7.4 Determiners of Singular and Uncount Nouns

Determiners with the code **singuncount** can only determine singular or uncount nouns. The examples in 67a. through 67c. show that the determiner *this* meets this criterion.

67a. This boy went home.
(count singular)
67b. This dirt was on the floor.
(uncount)
67c. *This boys went home.
(count plural)

68.
```
{base=this
    entry=E0060692
        cat=det
        variants=singuncount
        demonstrative
    }
```

4.7.5 Determiners of Plural and Uncount Nouns

Determiners with the code **pluruncount** can only determine count plural or uncount nouns. The examples in 69a. through 69c. show that *more* meets this criterion.

69a. More boys went home.
(count plural)
69b. More dirt was on the floor.
(uncount)
69c. *More boy went home.
(count singular)

70.
```
{base=more
    entry=E0040986
        cat=det
```

```
        variants=pluruncount
    }
```

4.7.6 Free Determiners

Determiners with the code **free** are not restricted as to the number of the nouns they can determine. The examples in 71a. through 71c. show that *some* is such a determiner.

71a. Some boys went home.
71b. Some dirt was on the floor.
71c. Some boy went home.

72.
```
{base=some
    entry=E0056728
        cat=det
        variants = free
    }
```

5. Complementation

5.1 Verb Complementation Patterns

The SPECIALIST lexicon recognizes five basic categories of verb, depending on the complements they take: intransitive, transitive, ditransitive, linking, and complex-transitive. The first four of these categories bear the names of the traditional verb types to which they correspond. Although the term "complement" is sometimes restricted to verb phrase constituents which follow linking verbs like *be*, or which enter into an "intensive relation" with an object of a verb, here a complement is considered to be any element of the verb phrase predicated by the verb. However, the particle of a verb particle construction is not treated as a complement of the verb. The code **part()** is added (following a semi-colon **;**) to the complementation codes to indicate a verb particle construction. See "Verb Particle Constructions" on page 48. Traditionally, the terms **intransitive**, **transitive**, and **ditransitive** refer to the number and type of noun phrases in the verb phrase. They are used here to refer to the number and type of complements in a verb phrase. **Linking** is a term traditionally used to refer to verbs which take a complement "referring to" or "in an intensive relationship" to another NP in the sentence. The term "linking" is retained and "complex-transitive" is used for verbs in which two complements may be regarded as being in a close semantic relationship with each other. Verbs can, and often do, fall into more than one category, and may have many complementation patterns within each category. The verb *give* illustrates the variety possible in complementation within a single verb entry.

73.
```
{base=give
    entry=E0029785
        cat=verb
        variants=irreg|give|gives|gave|given|giving|
        intran
        intran;part(out)
        intran;part(over)
        intran;part(in)
        intran;part(up)
        tran=np
```

tran=np;part(back)
tran=np;part(over)
tran=np;part(out)
tran=np;part(off);noprtmvt
tran=np;part(in)
tran=pphr(to,np);part(in)
tran=pphr(on,np);part(up)
tran=np;part(up)
tran=np;part(away)
tran=whinfcomp:arbc;part(away)
tran=ingcomp:subjc;part(up)
tran=ingcomp:subjc;part(out)
ditran=np,np|short shrift|
ditran=np,pphr(to,np);datmvt
ditran=np,pphr(to,np);part(back);datmvt
ditran=np,pphr(for,np)
ditran=np,pphr(to,np);part(out)
ditran=np,pphr(to,np);part(over)
ditran=np,pphr(for,np);part(up)
ditran=np,pphr(to,np);part(back)
cplxtran=np,infcomp:objr
cplxtran=np,infcomp:objr;part(out)
cplxtran=np,ingcomp:arbc;part(over)
nominalization=gift|noun|E0029737
}

The examples in 74a. through 74e. illustrate each major complementation category: 74a.

The bridge gave without warning. (intransitive)

74b. The handout gives the results. (transitive)

74c. He gave some money to charity. (ditransitive)

74d. He gave the Salvation Army his entire wardrobe. (ditransitive with dative movement)

74e. She was given to believe that her work was unacceptable. (complex transitive with object raising)

Note that several senses of the verb may be captured in the full entry for the verb.

5.1.1 Intransitive

A verb is marked intransitive if it may appear without any complement.

75a. He died.
intran

75b. The patient complained.
intran

Since adverbial particles are not considered complements, both *die* and *add* are intransitive in the following sentences:

76a. The patient gave up.

intran;part(up)

76b. Our fuel ran out.
intran;part(out)

Intransitive verbs may appear with non-complement adverbials.

77a. He died on Tuesday.
intran

77b. Our fuel ran out near Toledo.
intran

The intransitive verbs take no complement, but may take particles, as illustrated in Table 13.

Slot	Particle Code
intran	blank ;part(particle)[a]

Table 13. Fillers for the **intran**= slot

a. an particle (particle) code can be one of the following: *about, across, against, ahead, along, aloud, apart, around, at, away, back, behind, by, down, forth, forward, forwards, free, in it, off, on, open, out, over, round, through, to, together, up, without.*

5.1.2 Transitive

The slot **tran**= indicates a transitive verb. A verb is marked transitive if it takes a single complement. This may be a noun phrase, a prepositional phrase, or a clause. The codes following the equals sign refer to the syntactic category of the complement.

78a. I hit the boy. **tran=np**

78b. I propose that they initiate the project. **tran=fincomp(ts)**

78c. I propose going to the conference. **tran=ingcomp:arbc** 78d.
I asked whether he meant it. **tran=whfincomp**

The **tran**= slot takes ten different fillers as illustrated in Table 14.

Slot	Filler

tran=	np np|| pphr(,) binfcomp:interp infcomp:interp[a] ingcomp:interp ascomp:interp whinfcomp:interp fincomp() whfincomp

Table 14. Fillers for the **tran=** slot

a. an interpretation (interp) code can be one of **objc**, **subjc**, **arbc**, **nsc**.

Each of those slot fillers are discussed in "Verb Complements" on page 36.

An interpretation code (represented by *interp* in the table above) consists of one of the following: **objc**, **objr**, **subjc**, **subjr**, **arbc**, or **nsc.** Interpretation codes are discussed in "Interpretation Codes" on page 39.

5.1.3 Ditransitive

A **ditran**= slot indicates a ditransitive verb. A verb is marked ditransitive if it takes two complements, either of which might be a noun phrase, a prepositional phrase, or a clause.

79a. I struck Mary a blow.
ditran=np,np

79b. I wrote the letter to the president.
ditran=np,pphr(to,np);datmvt

79c. He increased the dose from 5 tablets to 6 tablets.
ditran=np,pphr(from,np,pphr(to,np))

79d. He showed us what he is able to do.
ditran=np,whfincomp

79e. They sided with him against the authorities.
ditran=pphr(with,np),pphr(against,np)

79f. The chairman must account to the stockholders for what he has done.
ditran=pphr(to,np),pphr(for,whfincomp)

The **ditran=** slot takes two fillers separated by a comma, SPECIALIST allows five different fillers for the first position and five for the second as illustrated in Table 15.

Slot	First filler	,	Second filler
ditran=	np np| | fincomp() ingcomp:interp pphr(,)		np whinfcomp:interp fincomp:interp fincomp() pphr(,)

Table 15. Fillers of the **ditran**= slot

The objects of ditransitive verbs may participate in Dative Movement. The SPECIALIST lexicon treats the dative shift phenomenon as a movement of the prepositionally marked indirect object into a position immediately following the verb. For example, "I gave Mary the book" is derived from "I gave the book to Mary'" by Dative Movement. The verb *give* is coded **ditran=np,pphr(to,np);datmvt**, with the additional code, **datmvt**, indicating that the two objects may be permuted. This device not only allows the two codes **ditran=np,pphr(P,np) and ditran=np,np1** to be collapsed into one code: **ditran(np,pphr(P,np));datmvt**, it captures the identity of the NPs in each code. The first NP of the **ditran(np,np)** code, the indirect object, is identified with the second NP in the **ditran=np,pphr(P,np),** and the second NP of the **ditran(np,np), the** direct object, is identified with the first NP of the **ditran=np,pphr(P,np)** code.

80a. I wrote a letter to the president.
80b. I wrote the president a letter.

5.1.4 Linking

The **link**= slot indicates a linking verb. A verb taking one complement which is in a close semantic relationship to the subject is a linking verb. The clearest example of this linking relation is shown with the auxiliary verb *be*, which "links" the subject to the object, but verbs such as *appear*, *feel*, *lie*, *remain*, *seem*, *become* and *get* are also included. Linking verbs may take a wide range of complements.

81a. He is **the chief**.
link=np
81b. He seems **to be a good boy**.
link=infcomp:subjr
81c. The joke is **in poor taste.**
link=advbl
81d. He is in New York.
link=advbl
81e. The problem is **believing everything he says**.
link=ingcomp:arbc
81f. The question is **where the scissors are now**.
link=whfincomp
81g. The director appeared **happy**.
link=adj
81h. He lives **in Silver Spring**.
link=advbl

So-called middle-verbs like *weigh* and *cost* are considered linking verbs.

82a. It weighs five pounds.
link=np

82b. It cost twenty dollars.
link=np

Eight fillers are allowed for the **link=** slot, as illustrated in Table 16.

Slot	Filler
link=	np pphr(,) adj advbl edcomp:interp infcomp:interp ingcomp:interp fincomp()

Table 16. Fillers of the **link=** slot

5.1.5 Complex-transitive

A **cplxtran**= slot indicates a complex transitive verb. A verb taking two complements which are in a close semantic relation to each other is a complex-transitive verb. The first complement is most often an NP, but it may also be a prepositional phrase, a finite clause, or an *-ing* clause. The second complement may be any one of a wide range of complements.

83a. I called him a fool.
cplxtran=np,np

83b. I consider him to be a genius.
cplxtran=np,infcomp:objr

83c. We can't risk him seeing us.
cplxtran=np,ingcomp:objr

83d. I regard him as my brother.
cplxtran=np,ascomp:objc

The **cplxtran** slot takes two fillers separated by a comma, as illustrated in Table 17.

Slot	First Filler	,	Second Filler
cplxtran=	np pphr(,) fincomp() ingcomp:interp		np adj advbl binfcomp:interp infcomp:interp ingcomp:interp ascomp:interp pphr(,) edcomp:interp

Table 17. Fillers of the **cplxtran=** slot

5.2 Verb Complements

Twelve different codes are used to characterize the complements for which the verb is subcategorized. (Note that in an actual lexical entry these are used in combination with one of the five basic transitivity categories for verbs.) Each complement type is illustrated below. In the examples, complements of the type being exemplified are emphasized (in **bold**).

5.2.1 Noun Phrase Complements (Objects)

The codes **np** and **np|N|** are used to indicate a noun phrase, i.e. a noun, its determiner and all of its pre- and post-modifiers.

84a. I hit **him**.
tran=np
84b. I saw **my best friend's favorite uncle, who owns the store**.
tran=np
84c. I gave **the book** to **my friend**.
ditran=np,pphr(to,np);datmvt
84d. I made **the decision**.
tran=np

When a verb idiomatically requires a particular noun as a complement, it is marked **np|N|**, where **N** stands for the required noun. Light verb constructions, however, like *take/have **a shower***, are not explicitly marked in this way, due to the plethora of possible noun complements. Instead they are covered by the verb complementation type, as in 84d above.

85a. take **account** of the situation
ditran=np|account|,pphr(of,np)
85b. shed **light** on
ditran=np|light|,pphr(on,np)

Wh-clauses introduced by *what* and *whatever*, sometimes called headless relatives, are considered to be a type of np. Other types of wh-clause complements are explicitly marked with either the code **whinfcomp** or **whfincomp.**

86a. I consider **what you do** impossible.
np

86b. I accept **what you said**.
np

5.2.2 Prepositional Phrase Complements

The codes **pphr(P,O)** and **pphr(P,O,pphr(P,O))** are used for a prepositional objects. P stands for the preposition and O for its object. The object may be one of several complement types. SPECIALIST does not assign verbs any higher level of transitivity than two (ditransitive). Sentences like *We decreased the dose from 5 mg. to 3 mg.* are analyzed as containing only one prepositional phrase in addition to the direct object NP. The *from...to* prepositional phrase is taken to have the structure reflected in the form of the code, **pphr(from,np,pphr(to,np))**. The code **pphr(P1,O,pphr(P2,O))** implies both **pphr(P1,O)** and **pphr(P2,O)** alone.

87a. The cavity filled **with fluid**.
tran=pphr(with,np)

87b. She knows **about their having suddenly left for New York**.
tran=pphr(about,ingcomp:arbc)

87c. The temperature increased **from ninety-five to one hundred**.
ditran=np,pphr(from,np,pphr(to,np))

88.
```
{base=increase
entry=E0034077
        cat=verb
        variants=reg
        intran
        tran=np
        tran=pphr(from,np)
        tran=pphr(from,np,pphr(to,np))
        tran=pphr(to,np)
        ditran=np,pphr(from,np)
        ditran=np,pphr(from,np,pphr(to,np))
        ditran=np,pphr(to,np)
    }
```

SPECIALIST allows nine fillers for the object position of pphr(_,_) as illustrated in Table 18.

Code	(	First Argument	,	Second Argument	)
pphr		Any Preposition		np np| | adj advbl infcomp:interp ingcomp:interp whinfcomp:interp binfcomp:interp edcomp:interp	

Table 18. Arguments of the pphr(,) code.

5.2.3 Adjective Complements

The code **adj** is used for an adjective, along with its modifiers and complements. Only linking and complex-transitive verbs take adjective complements.

89a. The moon appears **blue**.
link=adj
89b. The dinner seems **ready to eat**.
link=adj
89c. I dyed the cloth **red**.
cplxtran=np,adj
89d. I consider him **reckless**.
cplxtran=np,adj
89e. I painted the house **light green**.
cplxtran=np,adj
89f. The accused pleads **guilty**.
link=adj

5.2.4 Adverbial Complements

The code **advbl** is a cover term for simple adverbs and adverbial prepositional phrases. The latter is distinguished from the code **pphr(P,np)** where the particular preposition (P) is explicitly mentioned. An **advbl** prepositional phrase is not constrained to a particular preposition.

90a. He showed me **to the door**.
cplxtran=np,advbl
90b. He showed me **out**.
cplxtran=np,advbl

91a. She carried it **to the meeting**.

cplxtran=np,advbl

91b. She carried it **home**.
cplxtran=np,advbl

5.2.5 Non-Finite Clause Complements

The SPECIALIST lexicon provides for a variety of non-finite complements: infinitive clauses as in 92a. and 92b. (**infcomp** and **binfcomp**), present participle clauses as in 92c. (**ingcomp**), and past participle clauses as in 92d. (**edcomp**).

92a. He wants **to go home**.
92b. I helped **write the program**.
92c. The law forbids **stealing hubcaps**.
92d. He wants a new house **built**.

5.2.5.1 Interpretation Codes

Non-finite clauses generally do not have overt subjects; the understood subject of the clause is usually identified with some NP in the matrix sentence. Interpretation codes are provided to indicate where these logical subjects are found. The interpretation codes (**objc**, **objr**, **subjc**, **subjr**, **nsr**, **nsc**) are attached to the code for the non-finite complement with a colon. Each of them is discussed in sections 5.2.5.1.1 through 5.2.5.1.7 below.

5.2.5.1.1 Object Control

Object control means that the direct object in the higher clause is logically both the object of the higher verb and the subject of the embedded (non-finite) clause.

93a. I advised him **to go**.
cplxtran=np,infcomp:objc

93b. I told him **to rethink the problem**.
cplxtran=np,infcomp:objc

93c. We warned him against **stealing state secrets**.
cplxtran=np,pphr(against,ingcomp:objc)

Some object control sentences may be paraphrased with an embedded finite clause such that the object of the higher clause is coreferential with the subject of the lower clause. 94a. paraphrases 93a. and 94b. paraphrases 93b.

94a. I advised him that he should go.
94b. I told him that he should rethink the problem.

The interpretation code **objc** is attached (with a colon **:**) to complement codes which display object control.

5.2.5.1.2 Object Raising

Object raising means that the direct object in the higher clause is logically the subject of the non-finite clause and not the logical object of the higher clause.

95a. I intended him to win.
cplxtran=np,infcomp:objr
95b. I believe him to have stolen my watch.
cplxtran=np,infcomp:objr

Some object raising cases may be paraphrased with an embedded finite clause. 96a. is such a paraphrase of 95a. and 96b. paraphrases 95b.

96a. I intended that he would win.
96b. I believe that he stole my watch.

The interpretation code **objr** is attached (with a colon **:**) to complement codes indicating clauses which display object raising.

5.2.5.1.3 Subject Control

Subject Control means that subject of the higher clause is also the logical subject of the embedded infinitival clause:

97a. John promised to leap over the wall.
cplxtran=np,infcomp:subjc
97b. He is training to swim the channel.
tran=infcomp:subjc

Some subject control constructions may be paraphrased with an embedded definite clause having a subject coreferential with the subject of the higher clause. 98a. paraphrases 97a.

98a. John promised that he would leap over the wall.

The interpretation code **subjc** is attached (with a colon **:**) to complement codes indicating clauses which display subject control.

5.2.5.1.4 Subject Raising

Subject Raising indicates that the subject of the higher clause is the logical subject of the embedded infinitival clause:

99a. John seems to have criticized Martha.
link=infcomp:subjr
99b. John appears to firmly believe that he is Superman.
link=infcomp:subjr

Subject raising constructions may not be paraphrased in the same way that subject control constructions are:

100a. *John seems that he criticized Martha.

100b. *John appears that he firmly believes that he is Superman.

Expletive *it* subjects are possible in paraphrases of Subject Raising constructions:

101a. It seems that John has criticized Martha.
101b. It appears that John firmly believes that he is Superman.

The interpretation code **subjr** is attached (with a colon **:**) to the complement codes indicating clauses which display subject raising.

5.2.5.1.5 Arbitrary Control

Arbitrary control indicates that the subject of the lower clause is not linguistically controlled. 102a. can be paraphrased by 102b. and 102c. can be paraphrased by 102d. The subjects of *hunting* and *write* are not linguistically identified.

102a. I dislike hunting.
tran=ingcomp:arbc
102b. I dislike anyone hunting.

102c. I helped write the program.
tran=binfcomp:arbc
102d. I helped someone write the program.

The interpretation code **arbc** is attached (with a colon **:**) to complement codes indicating clauses which display arbitrary control.

5.2.5.1.6 Non-Subject Control

This phenomenon appears in present-participle clause complements of verbs. Non-subject control means that the subject of the matrix verb controls a missing non-subject NP in the present participle clause. The subject of the present participle clause is understood to be arbitrarily controlled. For example, 103a. can be roughly paraphrased by 103b. in which the object of *wash* is *the car*. The logical subject of *wash* is not linguistically indicated.

103a. This car needs washing.
tran=ingcomp:nsc
103b. This car needs someone to wash the car.
tran=ingcomp:nsc

103c. Surgical instruments require constant sharpening.
tran=ingcomp:nsc
103d. This abstract deserves careful reading.
tran=ingcomp:nsc

The code **nsr** is used to indicate non subject control. This phenomenon also appears in infinitive clause complements of adjectives. It is further discussed on page 55.

5.2.5.1.7 Non-Subject Raising

Non-subject raising is the phenomenon often called "tough-movement", wherein a non-subject noun phrase is missing from an infinitive complement and identified with an NP in the matrix clause. This phenomenon occurs primarily in infinitive complements of adjectives, and more rarely in adjective complements of nouns. We know of no verb which displays non subject raising. See page 50 for more detailed discussion.

5.2.5.2 Past Participle Clause Complements

The code **edcomp** is used for a past-participial clause. The code **edcomp** is required to have an interpretation code attached to it by a colon (**:**), indicating the subject of the past participial clause.

Table 19 illustrates the interpretation codes for **edcomp**.

Code	:	Interpretation code
edcomp		objc objr subjc subjr arbc

Table 19. Interpretation codes for edcomp

104a. I had the house **built by the best contractor in town**.
cplxtran=np,edcomp:objr;nopass
104b. The car got **washed**.
link=edcomp:subjc
104c. I saw the car **destroyed**.
cplxtran=np,binfcomp:objr

5.2.5.3 Infinitive Clause Complements

The code **infcomp** is used for an infinitive clause, introduced by *to*. The code **infcomp** must have an interpretation code.

105a. We can't afford **to do it**.
tran=infcomp:subjc
105b. He allowed the neighbors **to use the car**.
cplxtran=np,infcomp:objr
105c. He promised **to go**.
cplxtran=np,infcomp:subjc

The table below illustrates the interpretation codes for **infcomp**.

Code	:	Interpretation code
infcomp		objc objr subjc subjr arbc nsc

Table 20. Interpretation codes for infcomp in verb entries

An **infcomp** can have an overt subject if the complementizer *for* is present. This subject is not considered a complement of the verb but part of the to-infinitive complement. The interpretation code does not apply when there is an overt subject. The verb *afford* has the code **tran=infcomp:subjc** because, in the absence of an overt subject marked by *for*, the subject of the infinitive complement is understood to be the subject of the matrix sentence, as shown in 106a. In 106b. the occurrence of an overt subject (*the institution*) overrules the interpretation code.

106a. We can't afford **to make that commitment**.
tran=infcomp:subjc

106b. We can't afford **for the institution to make that commitment**.
tran=infcomp:subjc

5.2.5.4 Bare Infinitive Clause Complements

The code **binfcomp** is used for a "bare" infinitive clause, without *to*. The code **binfcomp** requires an interpretation code.

The table below illustrates interpretation codes for **binfcomp.**

Code	:	Interpretation code
binfcomp		objc objr subjc subjr arbc

Table 21. Interpretation codes for binfcomp in verb entries

107a. I had him **critique the documen**t.
cplxtran=np,binfcomp:objr

107b. Please help me **write the invitations**.

cplxtran=np,binfcomp:objc

5.2.5.5 Present Participle Complements

The code **ingcomp** is used for a present participial clause. The code **ingcomp** requires an interpretation code as illustrated in the following table.

Code	:	Interpretation code
ingcomp		objc objr subjc subjr arbc nsc

Table 22. Interpretation codes for ingcomp

108a. She advised **completing the project as soon as possible**.
tran=ingcomp:arbc

108b. I saw him **running down the street**.
cplxtran=np,ingcomp:objr

108c. I approve of him **finishing before the due date**.
cplxtran=pphr(of,np),ingcomp:objr

108d. This car needs washing.
tran=ingcomp:nsc

108e. We tried to prevent him from **dropping the course**.
cplxtran=np,pphr(from,ingcomp:objc)

An **ingcomp** can have a subject in the genitive case. As with the *for* subject of the to-infinitive, this subject is not considered a complement of the verb. Notice that the interpretation code does not apply when an explicit subject is present. The verb *report* has the code **tran=ingcomp:subjc** in both 109a. and 109b., indicating that the subject of the participial clause is controlled by the subject of *report*. But when there is an explicit genitive subject as in 109a. the interpretation code **subjc** is ignored.

109a. They reported **John's having noticed an error in the manuscript.**
tran=ingcomp:subjc

109b. They reported **having noticed an error in the manuscript**.
tran=ingcomp:subjc

5.2.6 Finite Clause Complements

Finite clause complements have tensed verbs, and show subject-verb agreement. They appear as full sentences generally introduced by the complementizer *that*. The code **fincomp()** is used to indicate finite clause complements.

5.2.6.1 Types of Finite Clause Complement

There are eight variations of **fincomp()**, each with or without an extraposed subject code, as illustrated in Table 23.

Code	(	Arguments	:	Extraposed Subject Code
fincomp		o t p s ts tp sp* tsp		subj

Table 23. Variations of fincomp( ). *Note: there are no current instances of the code sp in the lexicon.

Each of these codes is discussed in 5.2.6.1.1.- 5.2.6.1.8. below.

5.2.6.1.1 fincomp(o)

This code represents a finite clause with an optional *that* complementizer.

110a. They decided **(that) this was the right course of action**.
tran=fincomp(o)

110b. We assured him **(that) it would all turn out right**.
ditran=np,fincomp(o)

5.2.6.1.2 fincomp(t)

This code represents a finite clause with a required *that* complementizer.

111. I conclude **that you are a linguist**.
tran=fincomp(t)

5.2.6.1.3 fincomp(p)

This code represents a finite clause with an optional *that* complementizer which may be realized as the proform *so* or *not*.

112a. I assume **(that) you're going home tomorrow**.
tran=fincomp(p)
112b. I assume **so**.
tran=fincomp(p)

5.2.6.1.4 fincomp(s)

This code represents a subjunctive clause with an optional *that* complementizer. A subjunctive clause is a finite clause with an untensed verb (subjunctive form), or *should* followed by an untensed verb. Although finite complement was defined above in terms of having a tensed verb, subjunctive clauses are considered finite complements in spite of their apparent lack of tense.

113. I suggest **(that) you (should) be here on time**.
tran=fincomp(s)

5.2.6.1.5 fincomp(ts)

This code represents a subjunctive clause with a required *that* complementizer.

114. We require **that you (should) be here**.
tran=fincomp(ts)

5.2.6.1.6 fincomp(tp)

This code represents a finite clause with a required *that* complementizer, which may be realized as *so* or *not*.

115a. I advised him **that the cargo bay door was open**.
ditran=np,fincomp(tp)
115b. I advised him **so**.
ditran=np,fincomp(tp)
115c. I **so** advised him.
ditran=np,fincomp(tp)

5.2.6.1.7 fincomp(sp)

This code represents a subjunctive clause with an optional *that* complementizer, which may be realized as *so* or *not*. It is possible that such complements do not occur in English. The current SPECIALIST lexicon has no instances of **fincomp(sp).**

5.2.6.1.8 fincomp(tsp)

This code represents a subjunctive clause with a required *that* complementizer which can also be realized as *so* or *not*.

116. He moved **that this be stricken from the record**.

5.2.6.2 Extraposed Subject

The interpretation code **subj** is attached to a **fincomp( )** to indicate that the finite complement is an extraposed subject. That is, it appears postverbally only with an expletive *it* subject. For example, *appear* takes a **fincomp(p)** which is understood as the subject of *appear.* In 117a. and 117b. *so* and *that John won* are understood as the subject of *appear*.

117a. It appears **that John won**.
117b. It appears **so**.

118.
```
{base=appear
    entry=1
        cat=verb
        variants=reg
        intran
        tran=pphr(for,np)
        tran=fincomp(p):subj
        link=adj
        link=advbl
        link=infcomp:subjr
        link=np
}
```

Other examples of extraposed subject finite complements are illustrated in 119a. through 119d.

119a. It seems **that John intends to leave**.
tran=fincomp(o):subj
119b. It came about **that he lost the contest**.
tran=fincomp(t):subj;part(about);nopass
119c. It annoyed them **that John was late**.
ditran=np,fincomp(t):subj
119d.It seems **that John was the winner**.
tran=fincomp(p):subj

5.2.7 WH Complement Clauses

5.2.7.1 WH Finite Clause Complements

Finite clause complements introduced by a *wh*-word are coded **whfincomp**. Wh-finite complements can be introduced by *whether*, *how*, *why*, *where*, *when, who*, and sometimes *if*.

120a. He asked **whether they were coming**.
tran=whfincomp
120b. He asked me **if I'd be there on time.**
ditran=np,whfincomp
120c. I decided **how I would go about it**.

tran=whfincomp

120d. They expected to ascertain **whether this was appropriate**.
tran=whfincomp

5.2.7.2 WH Infinitive Complements

Infinitival clauses introduced by *wh*-words are coded **whinfcomp**; whinfcomp requires an interpretation code to indicate the subject. All **whinfcomps** in the SPECIALIST lexicon have been given the interpretation code **arbc**.

121a. I don't know **where to send the money**.
tran=whinfcomp:arbc

121b. I decided **how to go about it**.
tran=whinfcomp:arbc

121c. He chose **where to go**.
tran=whinfcomp:arbc

5.2.8 As Absolute Clause Complements

As absolute clauses consisting of a predicate introduced by *as* are coded **ascomp**. The predicate may be a noun, adverb, prepositional phrase or *ed*-clause. The code **ascomp** requires an interpretation code.

122a. We all looked upon him **as one of us**.
cplxtran=pphr(upon,np),ascomp:objr

122b. He qualified as **team chief**.
tran=ascomp:subjc

122c. The generals all regarded Uzbekistan **as well out of missile range**.
cplxtran=np,ascomp:objc

5.3 Verb Particle Constructions

English verbs often form a sort of compound with an adverbial particle. These particles themselves are entered in the lexicon as adverbs with a **modification_type=particle** code. Verbs which have particles, like *beat*, *tear*, and *run* have the code **part()** attached to the complemention codes for the complementation patterns in which the verb takes a particle.

123a. They **beat** him **up**.

123b. They **tore down** the theatre.

123c. They like to **run around**.

Each of the verb complement slots **intran**, **tran**=, **ditran**= and **cplxtran**= can be followed by a **part()** code indicating the particle of a verb particle construction. The argument of **part()** is the particle.

124. {base=tear

```
entry=1
    cat=noun
    variants=reg
entry=2
    cat=verb
    variants=irreg(tears,tore,torn,tearing)
    tran=np
    cplxtran=np,advbl
    cplxtran=np,advbl;part(away)
    cplxtran=np,advbl;part(off)
    cplxtran=np,advbl;part(out)
    cplxtran=np,advbl;part(up)
    cplxtran=np,pphr(off,np)
}
```

5.3.1 The Passive Construction

All of the complementation patterns represented in the SPECIALIST lexicon represent the verb phrase in active mood. Each transitive, ditransitive or complex-transitive code is subject to passivization unless it has been marked with the code **nopass**. The code **nopass** is added after a semi-colon (**;**) to those patterns which have no passive counterpart. The main verb *have* does not allow passive as illustrated in 125a. and 125b.

125a. He had a fancy coffee cup.
tran=np;nopass
125b. * A fancy coffee cup was had by him.

5.4 Noun Complementation

The slot **compl=** contains codes indicating noun complements. The possible complements of the noun are listed in Table 24.

Slot	Fillers
compl=	infcomp:interp fincomp() whfincomp whinfcomp:interp pphr(,) ascomp:interp

Table 24. Fillers of noun **compl=** slot

These codes have the same meaning as when they appear as complements of a verb.

5.4.1 Infinitive Clause Complements

Infinitive clause complements of nouns are coded with the code **infcomp**. Each **infcomp** is marked with an interpretation code; one of: **objr**, **objc**, **subjr**, **subjc**, **nsr**, or **arbc**. The interpretation codes are interpreted as they are for verbs, except that some nouns exhibit non-subject raising. (See "Infinitive Clause Complements" on page 44, and "Non Subject Raising" on page 52). The code **objc** indicates that the subject position of the **infcomp** may be controlled by an object genitive (indicated by the preposition *of,* or '*s* genitive). The code **subjc** indicates that the subject position of the **infcomp** can be controlled by a subjective genitive (*of* or '*s* genitive). The code **objr** indicates that the object genitive may be interpreted as the subject of the infcomp and not as an argument of the noun. The code **subjr** indicates that the subjective genitive may be interpreted as the subject of the infcomp and not as an argument of the noun. This is most clear in the case of nominalizations, where terms like subject and object can be analogized to the verb. Other abstract nouns can also have interpretation codes. For example, in the noun entry for *right* given in example 127, **infcomp:subjc** indicates that *John* and *suspects* in 126a. and 126b. are understood to be subjects of the infinitive complements and are arguments of *right* as well.

126a. John's right **to remain silent**.
126b. The right of suspects **to remain silent**.

127.
```
{base=right
    entry=E0053603
        cat=noun
        variants=reg
        variants=uncount
        variants=groupuncount
        compl=pphr(to,np)
        compl=infcomp:subjc
    }
```

5.4.2 Non Subject Raising

Nouns such as *breeze* and *snap* can trigger Non-Subject Raising. See "Non Subject Raising" on page 54 in the "Adjective Complementation" section.

The code **nsr** indicates that the subject of a predicate noun phrase is understood as some non-subject NP in its infinitive complement. The subject of the infinitive complement is arbitrarily controlled unless an explicit *for* subject is present.

128a. This pen is a breeze **to write with**.
128b. This tool is a snap **to use**.
128c. This tool will be a snap **for even the clumsiest linguist to use**.

129.
```
{base=breeze
    entry=E0014027
```

```
        cat=noun
        variants=reg
        compl=infcomp:nsr
    }
```

5.4.3 Finite Clause Complements

Finite clause complements for nouns are coded **fincomp**. As a noun complement **fincomp** is allowed the same range of variation as indicated for verbs in Table 22 on page 45.

130a. His awareness **that he was wasting time** was not sufficient.

131.
```
    {base=awareness
        entry=E0011455
            cat=noun
            variants=uncount
            compl=pphr(of,np)
            compl=fincomp(t)
            nominalization_of=aware|adj|E0011454
        }
```

5.4.4 WH Finite Clause Complements

The code **whfincomp** has the same meaning as a noun complement that it has as a verb complement. It represents a finite complement introduced by a *wh* word. See "WH Finite Clause Complements" on page 47 under Section 5.2.

132a. The reason **why he decided to attend** was unclear.

133.
```
    {base=reason
        entry=E0052127
            cat=noun
            variants=uncount
            variants=reg
            compl=infcomp:subjc
            compl=whfincomp
            compl=fincomp(t)
        }
```

5.4.5 Prepositional Phrase Complements

The prepositional phrase complement codes **pphr(P,0)** and **pphr(P,O,pphr(P,O))** have the same meaning in noun entries that they do in verb entries. See "Prepositional Phrase Complements" on page 37.

134a. The adequacy **of the railroad system for military transportation** was brought into question.

135. {base=adequacy
entry=E0007344
cat=noun
variants=uncount
compl=pphr(for,np)
compl=pphr(of,np)
nominalization_of=adequate|adj|E0007345
}

5.4.6 WH Infinitive Complements

This code is used for an infinitive clause, introduced by a *wh* word. As with verbs, **whinfcomp** must have an interpretation code. See "WH Infinitive Complements" on page 48.

136. The decision **when to leave** was hers alone.

137. {base=decision
entry=E0021039
cat=noun
variants=reg
variants=uncount
compl=infcomp:subjc
compl=fincomp(t)
compl=whinfcomp:arbc
compl=whfincomp
compl=pphr(of,np)
compl=pphr(by,np)
nominalization_of=decide|verb|E0021018
}

5.5 Adjective Complementation

Complements of adjectives are indicated as fillers of the **compl=** slot. The possible fillers for the complement slot of adjective entries are illustrated in Table 25.

Slot	Fillers
compl=	infcomp:interp fincomp( ) whfincomp whinfcomp advbl pphr(,) ascomp:interp

Table 25. Fillers of the adjective **compl**= slot.

5.5.1 Infinitive Clause Complements

The code **infcomp** indicates a *to* infinitive clause complement. As with verbs, interpretation codes attached with a colon are used to indicate to-infinitive complements. As an adjective complement infcomp can have **subjc**, **subjr**, **arbc**, **nsr**, or **nsc** attached. The codes **subjc**, **subjr** and **arbc** are used analogously to their use in verb complements. **objc** and **objr** do not apply to **infcomps** in adjective entries, since **compl**= slots do not take double arguments. Table 26 illustrates the interpretation codes that go with **infcomps** in adjective entries.

code	:	interpretation code
infcomp		subjc subjr arbc nsc nsr

Table 26. Interpretation codes for Adjectives

The noun phrase modified by the adjective, or the subject of the linking verb in a predicative construction, is considered the subject of the adjective. *They* is the subject of *eager* as well as the subject of *please* in 138a. and *problem* is the subject of *easy* as well as the subject of *solve* in 138b.

138a. They are eager **to please**.
compl=infcomp:subjc

138b. This is an easy problem **to solve**.
compl=infcomp:subjc

138c. They are careful **to avoid mistakes**.
compl=infcomp:subjc

138d. The author is ready **to write**.
compl=infcomp:subjc

138e. The potatoes are ready **to eat**.
compl=infcomp:nsc

The code **nsr** (for non subject raising) appears mainly with adjectives, and the code **nsc** (for non subject control) appears with **ingcomps** in verb entries and with **infcomps** in adjective entries.

```
139.        {base=eager
               entry=E0024307
                     cat=adj
                     variants=inv;periph
                     position=attrib(1)
                     position=pred
                     compl=infcomp:subjc
                     compl=fincomp(ts)
                     compl=pphr(about,np)
                     compl=pphr(for,np)
                     stative
                     nominalization=eagerness|noun|E0024309
               }
140.        {base=careful
               entry=E0015340
                     cat=adj
                     variants=inv;periph
                     position=attrib(1)
                     position=pred
                     compl=infcomp:subjc
                     compl=fincomp(t)
                     compl=whfincomp
                     nominalization=carefulness|noun|E0015342
}
```

5.5.2 Non Subject Raising

The code **nsr** indicates that the subject of the infinitive is arbitrarily controlled; that there is a non-subject NP missing from the infinitive clause; and that the NP which appears as the subject of the adjective is interpreted as this missing NP, not as the logical subject of the adjective. As example 141b. demonstrates, the missing NP need not be an argument of the verb of the infinitive clause. This phenomenon has been called Tough-Movement. Non-Subject Raising occurs only in adjectives and nouns.

141a. John is tough **to please**.
compl=infcomp:nsr
141b. This instrument is difficult **to teach students to fly with**.

142. {base=difficult
entry=E0022635
cat=adj
variants=inv;periph
position=attrib(1)
position=attribc
position=pred
compl=infcomp:nsr
stative
nominalization=difficulty|noun|E0022636
}

5.5.3 Non-Subject Control

The code **nsc** indicates that the subject of the infinitive clause is arbitrarily controlled; that there is a missing non-subject NP in the infinitive complement; and that the subject of the adjective is interpreted as the missing NP of the infinitive clause as well as the subject of the adjective.

143a. The book is ready **to print**.
compl=infcomp:nsc

The adjective *ready* participates in both non-subject control (infcomp:nsc) and subject control (infcomp:subjc), as evidenced by the ambiguity in example 144a.

144a. The chickens are ready **to eat**.
144b. The chickens are ready for someone to eat the chickens.
144c. The chickens are ready for the chickens to eat something**.**
144d. The chickens are ready.

Example 144b. is the non-subject control paraphrase of 144a. The subject control reading is paraphrased in 144c. Both readings of 144a. entail 144d. indicating that *the chickens* is the logical subject in both cases.

5.5.4 Finite Clause Complements

The code **fincomp( )** is used to indicate a finite clause complement as it is in entries for verbs. See “Finite Clause Complements” on page 45.

145. They were certain **that it would be possible**.

In some cases, the code **subj** is attached to a **fincomp( )** to indicate that the finite complement is an extraposed subject. That is, the finite complement can appear only after a predicate adjective with an expletive *it* subject.

146. It is likely that he will be there.
compl=fincomp(o):subj

147. {base=likely

```
entry=E0037555
        cat=adj
        variants=reg
        variants=inv;periph
        position=attrib(1)
        position=pred
        compl=infcomp:subjr
        compl=fincomp(o):subj
        stative
        nominalization=likelihood|noun|E0037554
}
```

5.5.5 WH Infinitive Clause Complements

As with verbs**, whinfcomp: interp** indicates an infinitive clause with a *wh*-element. It requires an interpretation code. See "WH Infinitive Complements" on page 48 in Section 5.2 .

148. He wasn't sure **whether to go**.

```
    compl=whinfcomp:subjc
```

149.

```
{base=sure
    entry=E0059287
        cat=adj
        variants=reg
        variants=inv
        position=attrib(1)
        position=attrib(3)
        position=pred
        compl=fincomp(o)
        compl=whfincomp
        compl=whinfcomp:subjc
        compl=infcomp:subjr
        compl=pphr(of,np)
        compl=pphr(of,ingcomp:subjc)
        stative
        nominalization=sureness|noun|E0059290
}
```

5.5.6 WH Finite Clause Complements

As with verbs, the code **whfincomp** represents a finite clause introduced by a *wh*-element. See "WH Finite Clause Complements" on page 47 in Section 5.2.

150. Be careful **what you say**.

```
    compl=whfincomp
```

5.5.7 Adverbial Complements

As with verbs, **advbl** indicates an adverbial complement, i.e. an adverb or prepositional phrase. See "Adverbial Complements" on page 38 in Section 5.2.

151. The building is situated **in a large field**.
compl=advbl

5.5.8 Prepositional Phrase Complements

The prepositional phrase codes **pphr(P,O)** and **pphr(P,O,pphr(P,O))** are used analogously to their use in verb entries. See "Prepositional Phrase Complements" on page 37.

152a. adequate **for the task**
152b. adequate **for solving the problem**
compl=pphr(for,np)

6. Nominalizations of Verbs and Adjectives

Verbs and adjectives which have nominalized forms have the slot **nominalization=** filled by the base form, category (noun) and EUI of the noun which is its nominalization. Nouns that are the nominalizations of verbs or adjectives have a **nominalization_of=** slot containing the base form, category and EUI of the verb or adjective of which they are the nominalizations. The elements of these cross references are separated by pipe "|" symbols. The verb *accumulate* has as its nominalization the noun *accumulation*. The noun phrase in 153b. is closely related in meaning to the sentence in 153a.

153a. John accumulated wealth.
153b. John's accumulation of wealth

154.
```
{base=accumulate
    entry=E0006764
        cat=verb
        variants=reg
        intran
        tran=np
        nominalization=accumulation|noun|E0006765
    }
```

155.
```
{base=accumulation
    entry=E0006765
        cat=noun
        variants=uncount
        variants=reg
        compl=pphr(of,np)
        compl=pphr(by,np)
```

nominalization_of=accumulate|verb|E0006764
}

7. Acronyms and Abbreviations

Acronyms and abbreviations have the slot **acronym_of=** or **abbreviation_of=** followed by the base form of their expansions. If the expansion is itself in the lexicon, the EUI of the expansion is included following a pipe "|" symbol.

156. {base=PCB
entry=E0004653
cat=noun
variants=metareg
acronym_of=polychlorinated biphenyl|E0048706
}

157. {base=polychlorinated biphenyl
entry=E0048706
cat=noun
variants=reg
variants=uncount
}

158. {base=ACh
entry=E0000061
cat=noun
variants=uncount
abbreviation_of=acetylcholine|E0006837
}

159. {base=acetylcholine
entry=E0006837
cat=noun
variants=reg
variants=uncount
}

8. Proper Nouns

Proper nouns are marked with the feature **proper**.

160. {base=Austria
entry=E0000523

cat=noun
variants=uncount
proper
}

9. Adjective Positions

SPECIALIST recognizes two main positions for adjectives: attributive and predicative. An adjective between the determiner and the head noun of a noun phrase is in attributive position. One following the verb either as a predicate adjective in a linking construction or as the second member of a complex transitive construction is predicative. Most adjectives are both attributive and predicative. In addition, SPECIALIST recognizes attributive adjectives with discontinuous complements, and adjectives that post-modify nouns. The **position=** slot records position information for adjectives.

The possible fillers for the position slots of adjectives are listed below:

Slot	Fillers
position=	attrib(1) attrib(2) attrib(3) attribc pred post

Table 26. Fillers of the **position=** slot

9.1 Attributive Adjectives

Attributive adjectives occur between determiners and nouns:

161. The **red** apple was eaten.

The code **attrib(N)** appears in the **position**= slot of attributive adjectives; **N** is a number (1-3) to indicate its place in the normal sequence of adjectives. Some adjectives can only occur in attributive positions, e.g. *mock*, *favorite* and *roast*. But most adjectives occur in both predicative and attributive positions. As the examples in 162a. through 162d. illustrate, *roast* is attributive only and *raw* is both attributive and predicative.

162a. They ate **roast** lamb for dinner.
162b. *The lamb they ate for dinner was **roast**.
162c. They ate **raw** vegetables for dinner.
162d. The vegetables they ate for dinner were **raw**.

9.1.1 Position Classes for Attributive Adjectives

Each attributive adjective is assigned a number indicating its type. The numbers reflect the typical order of occurrence in an NP. Qualitative adjectives (attrib(1)) generally precede color adjectives (attrib(2)) and classifying adjectives (attrib(3)). Color adjectives generally precede classifying adjectives.

163a. The **big red stone** house.
attrib(1), attrib(2)
attrib(3)
163b. ? The **red stone big** house.
attrib(2),attrib(3),attrib(1)

Acute is qualitative and classifying in different senses. When *acute* means perceptive, as in example 164a., it is qualitative (attrib(1)); when it means "having a short and relatively severe course" as in example 164b., it is classifying (attrib(3)).

164a. It was apparent only to the most **acute** observer.
position=attrib(1);
164b. He suffers from **acute abdominal** hemorrhage.
164c. ? He suffers from **abdominal acute** hemorrhage.

Abdominal is a classifying adjective. Example 164c. is unnatural because qualitative adjectives generally precede classifying adjectives.

165.
```
{base=acute
    entry=E0007127
        cat=adj
        variants=inv;periph
        position=attrib(1)
        position=attrib(3)
        position=pred
        stative
        nominalization=acuity|noun|E0007121
        nominalization=acuteness|noun|E0007129
    }
```

166.
```
{base=abdominal
    entry=E0006444
        cat=adj
        variants=inv
        position=attrib(3)
        position=pred
        stative
    }
```

Collin's COBUILD Dictionary makes these distinctions and has been consulted in coding adjectives. Each of the position classes is described below in sections 9.1.1.1 through 9.1.1.3.

9.1.1.1 Qualitative Adjectives

A qualitative adjective can have more or less of the quality it describes, and it may be preceded by an intensifier. If several adjectives are used together, qualitative adjectives come before color and classifying adjectives. The code **attrib(1)** appears in the **position=** slot of qualitative adjectives. *Severe* is a qualitative adjective.

167a. A **severe** burn
167b. The most **severe** burn
167c. A **severe abdominal** pain
167d. ? An **abdominal severe** pain

168.
```
{base=severe
    entry=E0055474
        cat=adj
        variants=reg
        position=attrib(1)
        position=pred
        nominalization=severeness|noun|E0055476
        nominalization=severity|noun|E0055478
    }
```

9.1.1.2 Color Adjectives

A color adjective may be modified by an adjective which describes the color more exactly, such as *clear*, *pale*, or *bright*. In a group of adjectives, the color adjective comes after any qualitative and before any classifying adjectives. The code **attrib(2**) appears in the **position=** slot of color adjectives.

169a. A rather unusual **red** flower
169b. An old **brown** wooden desk
169c. The **purple** gown

170.
```
{base=purple
    entry=E0051252
        cat=adj
        variants=reg
        position=attrib(2)
        position=pred
        stative
        nominalization=purpleness|noun|E0337506
    }
```

9.1.1.3 Classifying Adjectives

A classifying adjective cannot have more or less of the quality it describes, unlike qualitative adjectives. In a string of adjectives, classifying adjectives follow qualitative and color adjectives and come closest to the noun. The code **attrib(3)** appears in the **position=** slot of classifying adjectives. Many technical adjectives like *cardiac* which are closely related to nouns are classifying.

171a. A **steel grey cardiac** monitor
171b. ? A **cardiac steel grey** monitor.
171c. *The most **cardiac** monitor.

172.
```
{base=cardiac
    entry=E0015228
        cat=adj
        variants=inv
        position=attrib(3)
        position=pred
        stative
    }
```

Adjectives denoting materials like *woollen*, *silken* and *earthen* are usually classifying. *Wooden*, in the sense "made of wood" is a typical classifying adjective. In the sense "leaden, stiff, unemotional" it is qualitative.

172a. A **brown wooden** desk
172b. ? A **wooden brown** desk
172c. * A more **wooden** desk.

173.
```
{base=wooden
    entry=E0065576
        cat=adj
        variants=inv
        position=attrib(1)
        position=attrib(3)
        position=pred
        stative
        nominalization=woodenness|noun|E0065577
    }
```

9.1.2 Attributive Adjectives with Complements

"Attributive complement" adjectives are those attributive adjectives which have a discontinuous complement occurring after the noun. The code **attribc** indicates an attributive adjective which may take a complement.

174a. It was an **easy** problem **to solve**.

174b. He is a **difficult** person **to talk to**.

Non Subject Raising ("tough movement") adjectives are generally **attribc** but not all **attribc** adjectives participate in Non Subject Raising.

9.2 Predicative Adjectives

Predicative adjectives occur after the verb and are generally in a close semantic relation to the subject, as in 175a. or in a semantic relation to another object complement, as in 175b.

175a. The painting is **ugly**.
175b. He thought the painting **ugly**.

Predicative adjectives may have complements.

176a.He is **happy** to be here.
176b. He is **hopeful** that he will win.
176c. He was **slow** getting up.
176d. The piano is **impossible** for me to move.
176e. The problem is **easy** to solve.

Most adjectives can occur predicatively. Adjectives which can occur predicatively have the filler **pred** in their **position=** slot.

9.3 Post-Nominal Adjectives

The code **post** appears in the **position**= slot of post-nominal attributive adjectives. Post-nominal attributive adjectives occur after the nouns they modify:

177a. The man **responsible** went to prison. 177b.
The president **elect** campaigned hard.

Post-nominal adjectives are quite rare. Some like *elect* and *galore* occur only post-nominally.

177c. The president **elect**
177d. *The **elect** president
177e. presents **galore**
177f. ***galore** presents

178.
```
{base=elect
    entry=E0024672
        cat=adj
        variants=inv
        position=post
        stative
    }
```

In other cases there is a sense difference which depends on whether a form is prenominal or post-nominal. *Proper* is attributive in the sense shown in 179a. It is post-nominal in the sense shown in 179b.

179a. **proper** writing
179b. writing **proper** (excluding say rebuses and pictograms)

180.
```
{base=proper
    entry=E0050395
        cat=adj
        variants=inv
        variants=inv;periph
        position=attrib(1)
        position=attrib(3)
        position=pred
        position=post
        compl=fincomp(ts):subj
        compl=pphr(to,np)
}
```

Post-nominal adjectives include culinary (often borrowed) adjectives like: *florentine*, *carbonara*, and *newburg*.

The **post** code is reserved for those few adjectives which either must occur post-nominally or occur post-nominally in a special sense. Post-nominal occurrences of ordinary adjectives are considered to involve a reduced relative clause. The examples in 181a. through 181c. are not considered instances of post-nominal adjectives.

181a. a sword **red** with blood
181b. a friend **worried** about his health
181c. an officer of the court **sworn** to uphold the law

One symptom of the difference between true post-nominal adjectives and adjectives in a reduced relative construction is that ordinary adjectives need to be conjoined or have complements in a reduced relative construction.

182a. *a sword **red**
182b. *a friend **worried**
182c. *an officer of the court **sworn**

10. Stative Adjectives

The feature **stative** is added to all adjectives which are static, or relatively unchanging, in nature. If an adjective cannot be used with the progressive or imperative, it is considered stative. The adjective *tall* is stative, as 183b. and 183c. show. *Nasty* is not stative, as 184b. and 184c. show.

183a. He is tall.
183b. *He is being tall.
183c. *Be tall.

184a. He is nasty.
184b. He is being nasty.
184c. Be nasty.

185.
```
{base=tall
    entry=E0059841
        cat=adj
        variants=reg
        position=attrib(1)
        position=pred
        position=post
        stative
        nominalization=tallness|noun|E0059842
    }
```

186.
```
{base=nasty
    entry=E0041957
        cat=adj
        variants=reg
        position=attrib(1)
        position=pred
        position=attribc
        compl=infcomp:arbc
        nominalization=nastiness|noun|E0041956
    }
```

11. Adverb Modification types.

Adverbs are marked in SPECIALIST to indicate their modification type. Each adverb entry must have at least one **modification_type=** slot. This slot takes four different fillers. Sentence modifiers and verb modifiers have a **manner**, **temporal** or **locative** feature. Table 28 illustrates the fillers of the **modification_type**= slot.

Slot	**Fillers**	**;**	**Features**
modification_type=	sentence_modifier verb_modifier		manner temporal locative
	particle intensifier		

Table 28. Fillers of the **modification _type**= slot

11.1 Adverbial Particles

Particles are listed as adverbs in the SPECIALIST lexicon. They are given the modification type **particle**.

187. He backed **up** the hard disk.
modification_type=particle

188.
```
{base=up
    entry=E0063424
        cat=adv
        variants=inv
        modification_type=verb_modifier;locative
        modification_type=particle
    }
```

11.2 Intensifiers

Intensifiers modify adjectives or adverbs. They are marked with the modification type **intensifier**.

189. They are **quite** happily married.

190.
```
{base=quite
    entry=E0051682
        cat=adv
        variants=inv
        modification_type=intensifier
        modification_type=verb_modifier;manner
        }
```

11.3 Sentence Modifiers

Sentence modifiers modify whole sentences. The code for sentence modifiers is **sentence_modifier**.

191a. **Unfortunately**, I wasn't able to visit him. 191b. **Frankly**, it wasn't very good.

192.
```
{base=frankly
    entry=E0028745
        cat=adv
        variants=inv;periph
        modification_type=intensifier
        modification_type=sentence_modifier;manner
        modification_type=verb_modifier;manner
}
```

11.4 Verb Modifiers

Verb modifiers modify the verb phrase. The code for verb modifiers is **verb_modifier**.

193a. Fred typed **busily** all day.

193b. He worked **slowly** and **carefully.**

194. {base=carefully
entry=E0015341
cat=adv
variants=inv;periph
modification_type=verb_modifier;manner
}

11.5 Locative, Temporal and Manner Adverbs

Sentence and verb modifiers are given one additional feature from the set **locative**, **temporal**, or **manner**.

11.5.1 Locative

Locative adverbs indicate direction, location, etc. As a diagnostic, the interrogative adverb *where* generally elicits a locative adverbial answer.

195a. He didn't run **far**.
195b. He walked the dog **around**.

196. {base=around
entry=E0010426
cat=adv
variants=inv
modification_type=verb_modifier;locative
modification_type=particle
}

11.5.2 Temporal

Temporal adverbs indicate time, duration, etc. As a diagnostic, the interrogative adverb *when* generally elicits a temporal adverbial answer.

197a. The newspaper arrives **daily**.
197b. She exercises **often**.

198. {base=often
entry=E0043653
cat=adv
variants=reg
modification_type=verb_modifier;temporal
}

11.5.3 Manner

Manner adverbs indicate the way an action was accomplished. As a diagnostic, the interrogative adverb *how* generally elicits a manner adverbial answer.

199a. The teacher **politely** assessed the student. 199b.
They taunted him **cruelly**.

200. {base=politely

entry=E0048641
cat=adv
variants=inv;periph
modification_type=verb_modifier;manner
}

12. Interrogative

Certain pronouns, adverbs and determiners are subject to wh-movement in the formation of questions, and relative clauses. These wh-elements are marked with the code **interrogative**.

12.1 Interrogative Pronouns

The interrogative pronouns are: *what*, *whatever*, *when*, *where*, *which*, *whichever*, *who*, *whoever*, *whom*, *whomever*, and *whose*.

201a. What did you say?
201b. Who did you see?

202. {base=what
entry=E0065275
cat=pron
variants=third
gender=neut
interrogative
type=subj
type=obj
}

12.2 Interrogative Adverbs

The interrogative adverbs are: *how*, *however*, *when*, *whence*, *whenever*, *where*, *wherein*, *wherever*, *whither*, *why*, *whyever*.

203a. How did you do that?
203b. Where have you been?

204. {base=why
entry=E0065412
cat=adv
variants=inv
interrogative
modification_type=verb_modifier;manner
}

12.3 Interrogative Determiners

The interrogative determiners are: *what*, *whatever*, *which*, *whichever*.

205a. Which book did you buy?
205b. What book would you recomend?

206.
```
{base=which
    entry=E0065318
        cat=det
        variants=free
        interrogative
    }
```

13. Negation

There are two classes of negative adverbs recognized in SPECIALIST, true sentence negators and broadly negative adverbs. The two types are discussed in 13.1 and 13.2.

13.1 True Negative Adverbs

Negative adverbs, which produce true sentence negation, are given the feature **negative**. Sentence negation produces a sentence contradictory to its positive counterpart. 207b. is a sentence negation of 207a. as evidenced by the fact that 207c. is contradictory.

207a. I eat pizza. 207b.
I never eat pizza.
207c. * I eat pizza and I never eat pizza.

208.
```
{base=never
    entry=E0042565
        cat=adv
        variants=inv
        modification_type=sentence_modifier;temporal
        negative
    }
```

The negative contracted forms of modals and auxiliaries also have the feature **negative** with the same meaning.

13.2 Broadly Negative Adverbs

The feature **broad_negative** is added to the entries of terms like *hardly*, *scarcely*, *rarely*, *barely* and *seldom* which are broadly negative.

Broadly negative adverbs are not strictly sentence negators, but they trigger certain syntactic phenomena associated with negation, specifically polarity items, positive question tags, and fronting with subject/aux inversion. They are not strict sentence negators in that they do not produce a contradiction when conjoined to their positive counterpart. 209a. and 209b. are not contradictory because *hardly* and *seldom* are broad negative adverbs. Compare those to 210a. and 210b. which are contradictory.

209a. I hardly finished the work, but I did finish it.
209b. He seldom fails exams, but sometimes he does.

210a. *He never fails exam but sometimes he does.
210b. *He did not finish the work, but he did finish it.

Broadly negative adverbs participate in syntactic phenomena associated with negation. Three such phenomena are discussed below.

Negative polarity items:

Broadly negative adverbs allow negative polarity items, just as strict negatives do.

211a. He never lifted a finger to help. (strict negative)
211b. I have hardly any beans left. (broad negative)
211c. I seldom had any beans left. (broad negative)
211d. *I have any beans left.

212a. He did not lift a finger to help. (strict negative)

213a. He hardly lifted a finger to help. (broad negative)
213b. He seldom lifted a finger to help. (broad negative)
213c. *He lifted a finger to help.

Positive question tags:

Strict negatives co-occur with positive question tags. The broad negative adverb *scarcely* in 214b. and 214d. is subject to dialectal variation - for some speakers it is grammatical only with positive question tags, like the strict negation in 214a. and 214c., but for other speakers it is grammatical only with negative question tags, as in 214d. and 215b.

214a. They don't seem to care, do they? (strict negative)
214b. (*)They scarcely seem to care, do they? (broad negative)
214c. *They don't seem to care, don't they? (strict negative)
214d. (*)They scarcely seem to care, don't they? (broad negative)

215a. (*)They hardly have any friends, have they?
215b. (*)They hardly have any friends, haven't they?

Fronting with subject/aux inversion:

Both strict and broad negatives allow subject/aux inversion with fronting. The broad negative in 216b. allows inversion with fronting just like the strict negative in 216a.

216a. Never had he seen a more perfect apple. (strict negation) 216b.
Hardly have I ever used a better dictionary. (broad negative)

217.
```
{base=seldom
    entry=E0055050
        cat=adv
        variants=inv
        modification_type=sentence_modifier;manner
        broad_negative
    }
```

14. Pronouns

Pronouns in the SPECIALIST lexicon are involved in 3 systems:

1. variants - concerns person/number verb agreement.
2. gender - concerns reference/coreference.
3. type - concerns government, possession, reflexiveness, quantification and deixis.

Each of these systems is discussed in sections 14.1 through 14.3 below.

14.1 Person and Number

Person and number are matters of agreement or concord. The codes which mark person and number for pronouns are recorded as fillers of the **variants=** slot. They are described in Section 4.6 on page 26. Table 11, which describes the fillers of the **variants**= slot for pronouns, is repeated here for convenience.

Slot	Fillers
variants=	fst_plur fst_sing free sec_plur sec_sing second third thr_plur

Table 29. Fillers of the **variants**= slot for pronouns.

14.2 Gender for Personal Pronouns

The **gender**= slot records restrictions on the referents/antecedents of pronouns in terms of humanness and sex.

Slot	filler
gender=	pers(masc) pers(fem) pers neut

Table 30. Fillers of the **gender**= slot for pronouns

The organization of gender codes in SPECIALIST is illustrated in Figure 1 on page 73. Pronouns marked **pers** refer only to humans (or figurative humans), **pers(masc)** pronouns refer only to male humans (or figurative male humans) and **pers(fem)** pronouns refer to female humans (or

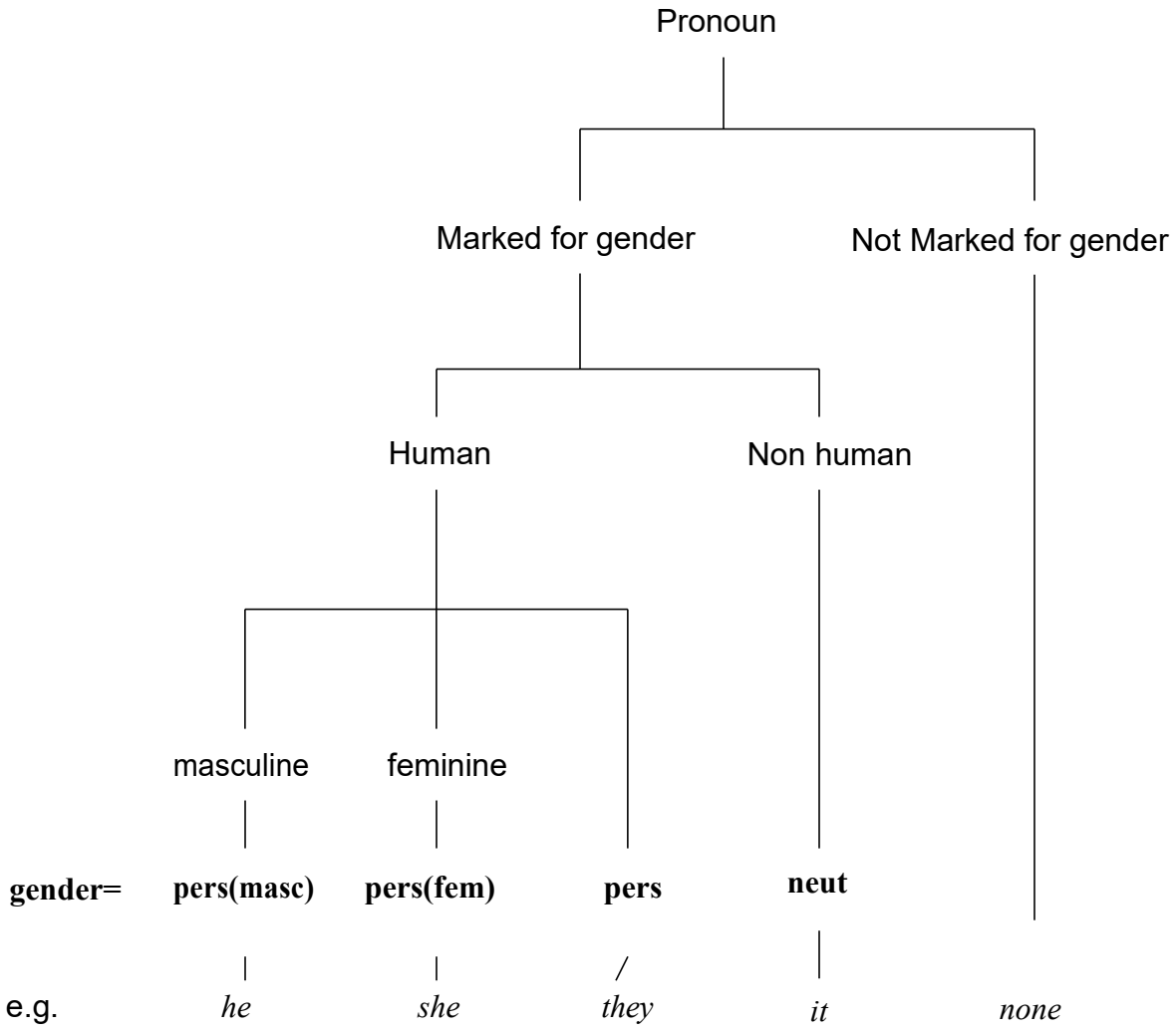

Figure 1. gender codes for pronouns

figurative female humans). Pronouns marked **neut** (for neuter) refer only to non-humans. Pronouns not marked for gender place no human or sex restrictions on their referents or antecedents.

218. {base=she
entry=E0055585
cat=pron
variants=thr_sing
gender=pers(fem)
type=subj

```
                }
219.            {base=he
                    entry=E0030918
                        cat=pron
                        variants=thr_sing
                        gender=pers(masc)
                        type=subj
                    }
220.            {base=it
                    entry=E0036100
                        cat=pron
                        variants=thr_sing
                        gender=neut
                        type=subj
                        type=obj
                    }
```

14.3 Type

Type covers a variety of phenomena, discussed in 14.3.1 through 14.3.6. The fillers of the **type=** slot are summarized in Table 31.

slot	filler	Concerned with
type=	obj subj	Government
	poss possnom	Possession
	refl	Reflexivity
	univ indef(neg) indef(assert) indef(nonassert)	Quantification
	dem	Deixis

Table 31. Fillers of the **type=** slot for pronouns

14.3.1 Government

Each pronoun has a type code indicating its case: **subj** for nominative case and **obj** for accusative case. Nominative case pronouns can occur in subject position and accusative case pronouns can

occur in object position. The pronoun *he* is nominative (**subj**), not accusative (**obj**); so 221b. is ill formed. *Him* is accusative (**obj**), not nominative (**subj**); so 222b. is ill formed.

221a. He talked to the committee.
221b. * The committee talked to he.

222a. The committee talked to him.
222b. * Him talked to the committee.

223.
```
{base=he
    entry=E0030918
        cat=pron
        variants=thr_sing
        gender=pers(masc)
        type=subj
    }
```

224.
```
{base=him
    entry=E0031685
        cat=pron
        variants=thr_sing
        gender=pers(masc)
        type=obj
    }
```

Pronouns which can occur either in subject or object positions get both **subj** and **obj** type codes. The pronoun *it* can be either nominative or accusative.

225a. It saw me.
225b. I saw it.

226.
```
{base=it
    entry=E0036100
        cat=pron
        variants=thr_sing
        gender=neut
        type=subj
        type=obj
}
```

14.3.2 Possession

The type codes **poss** and **possnom** indicate possessive pronouns. Pronouns marked **poss** are pre-nominal possessive pronouns which occur before nouns in noun phrases; they cannot occur as noun phrases or heads of noun phrases.

227a. My book is on the table.

227b. His book is on the table.
227c. Its name is Fido.
227d. *My is on the table.

The pronouns marked **poss** are: *her, his, its, my, one's, our, their, whose and your.*

228.
```
{base=your
    entry=E0065895
        cat=pron
        variants=second
        gender=pers
        type=poss
    }
```

Pronouns marked **possnom** are possessive nominals which can be noun phrases, or appear in predicative constructions.

229a. Mine is on the table.
229b. That book is mine.
229c. Yours died.
229d. The tadpole that died was yours.
229e. *Her died. (poss)
229f. *The tadpole that died was her. (poss)

Possessive nominals are doubly pronominal*; mine* represents the second person and refers to something possessed by the second person. The possessive nominal pronouns are: *hers, his, mine, ours, theirs,* and *yours.*

230.
```
{base=yours
    entry=E0065896
        cat=pron
        variants=second
        gender=pers
        type=possnom
    }
```

Two pronouns, *his* and *whose*, are both pre-nominal possessives and possessive nominals. They are given both codes **poss** and **possnom**. Possessive (genitive) noun phrases are like *his* in this respect.

231a. His book is on the table.
231b. His is on the table.
231c. The book on the table is his.

232a. John's book is on the table.
232b. John's is on the table.
232c. The book on the table is John's.

233. {base=his
entry=E0031715
cat=pron
variants=thr_sing
gender=pers(masc)
type=poss
type=possnom
}

14.3.3 Reflexive

Reflexive pronouns, which occur obligatorily in the same clause following their antecedents, are marked with the type code **refl**.

234a. John saw himself.
234b. *Himself saw John.

The reflexive pronouns are: *herself, himself, itself, myself, oneself, themself, themselves, yourself*, and *yourselves.*

235. {base=herself
entry=E0031461
cat=pron
variants=thr_sing
gender=pers(fem)
type=refl
type=obj
}

14.3.4 Quantification

Pronouns which have quantifier like properties receive one of the type codes **univ, indef(assert)**, **indef(nonassert)** or **indef(neg)**.

Pronouns which are manifestations of a universal quantifier are marked with the code **univ** in their **type=** slot. The universally quantified pronouns are: *all*, *each*, *everybody*, *everyone*, *everything.*

Indefinite pronouns are divided into three series corresponding to the quantifiers they contain; *some*, *any* or *no*. Indefinite pronouns of the non-assertive series, (*any*) are marked **indef(nonassert)**; indefinite pronouns of the assertive series (*some*) are marked **indef(assert)**; and indefinite pronouns of the negative series (*no*) are marked **indef(neg)**. The code **indef(neg)** is equivalent to the feature **negative**. It indicates a sentence negation.

1. The **indef(nonassert)** pronouns are: *any, anybody, anyone,* and *anything.*

236. {base=anybody
entry=E0009846

cat=pron
variants=thr_sing
gender=pers
type=indef(nonassert)
type=obj
type=subj
}

2. The **indef(assert)** pronouns are: *some, somebody, someone,* and *something.*

237. {base=something
entry=E0056737
cat=pron
variants=thr_sing
gender=neut
type=indef(assert)
type=obj
type=subj
}

3. The **indef(neg)** pronouns are: *nobody, none, no one*, and *nothing*.

238. {base=none
entry=E0042838
cat=pron
variants=third
type=indef(neg)
type=obj
type=subj
}

14.3.5 Deixis

Demonstrative pronouns have the code **dem** in their **type=** slot. The demonstrative pronouns are: *that, these, this* and *those.* These pronouns are also demonstrative determiners, and the **type=dem** code is equivalent to the **demonstrative** code for determiners.

239a. That is a big fish.
239b. This is an even bigger one.

240. {base=this
entry=E0060693
cat=pron
variants=thr_sing
type=dem
type=obj

type=subj
}

14.3.6 Deictic Determiners

The feature **demonstrative** marks the deictic determiners: *this*, *that*, *these*, and *those*. Each of those determiners is also a demonstrative pronoun.

241a. That fish is really big. 241b.
This fish is even bigger.

242. {base=that
entry=E0060479
cat=det
variants=singuncount
demonstrative
}

15. Classification types

The SPECIALIST lexicon also marks various usage-related features of lexical variants. These features are marked by five different fillers of the **class_type** slot:

1. **archaic** - indicates that a variant is no longer in common use
2. **informal** - indicates that the variant is used primarily in colloquial contexts
3. **source** - indicates the language or dialect where a variant originated
4. **taxonomic** - indicates that a variant is a term from biological taxonomy (genus, species, etc)
5. **other** - indicates some other type of classification information (gene, protein, etc)

Each of these **class_type**s is discussed in sections 15.1 through 15.5 below.

15.1 Archaic variants

Lexical entries that contain variants with **class_type**=**archaic** are those with a variant no longer used as a common form of a lexical item. These terms may have more modern equivalents within the same entry ("cold" as a synonym of the archaic "colde") or in separate lexical entries entirely ("the" as a synonym of the archaic "ye"). In either case, only the spelling variants which are archaic are tagged as such, using the following syntax: **class_type=archaic**|{archaic base form}

243. {base=cold
spelling_variant=colde

```
entry=E0017762
    cat=noun
    variants=reg
    variants=uncount
    class_type=archaic|colde
    }
```

15.2 Informal variants

Most entries in the SPECIALIST lexicon have been sourced from written work, so they are primarily used in formal (academic, published) contexts. The **class_type**=**informal** variants are those lexical items which are used instead in primarily informal contexts. That is, they may be generally seen only in spoken language, or in personal notes or shorthand rather than in published documents. Colloquial terms like those seen in internet forum posts, online chats or text messaging are also included in **class_type=informal**. Often these variants will be shortened versions of more standard/formal variants ("u" as a variant of "you"; "pls" as a variant of "please", "Xmas" as a variant of "Christmas") but this is not always the case.

Informal tags come with synonym information; the syntax is as follows: **class_type=informal**|{informal base form}|{citation form of the synonym}{entry number of the synonym, if available}
When informal synonyms are not lexicon entries, then the final field is excluded. These synonyms are unidirectional synonyms, meaning that the formal equivalent may be a superset of the informal one. However, the part of speech of informal base forms and synonyms must always be the same. In the example below, "nite" is an informal variant (and a spelling variant) in the entry for "night", so it is tagged as a synonym of itself. If an informal term has multiple distinct senses, this information will be given by multiple **class_type** tags.

```
244.        {base=night
              spelling_variant=nite
              entry=E0042638
                  cat=noun
                  variants=reg
                  variants=uncount
                  compl = pphr(of, np)
                  class_type=informal|nite|night|0042638
                  }
```

15.3 Source variants

One of the main sources of spelling variation in the Lexicon comes from the distinction between British English and American (US) English. The **class_type**=**source** tag allows for the

classification of spelling variants as British English variants. Note that although the language attribute is termed 'British', Australia and Canada often use British English variants rather than American ones. If there are terms used exclusively in e.g. Australia and not Britain, this will be specified in the annotation slot.

Also included in this tag are foreign-language borrowings into English. The source of these terms is therefore listed as the language of origin for the borrowing. Options for the **source** are: {latin, greek, french, spanish, italian, german, british, canadian, other}. When the language of origin is not included in the list, the filler **other** is used, and the language origin information is given in the annotations slot. The syntax for source information is as follows:
class_type=source|{source variant}|{source language option}

245.
```
{base=color
    spelling_variant=colour
    entry=E0017902
        cat=noun
        variants=reg
        variants=uncount
        class_type=source|colour|british
        }
```

246.
```
{base=bonafide
    entry=E0419799
        cat=adv
        variants=inv
        modification_type=intensifier
        class_type=source|bonafide|latin
        }
```

15.4 Taxonomic variants

Lexical entries that contain variants with **class_type=taxonomic** are those with a variant that is a term from biological taxonomy. These terms may be combinations of genus and species terms (e.g. “Homo sapiens”), abbreviations of such (“H. sapiens”) or other taxonomic subdivisions, like families or clades. Most typically these lexical variants are capitalized, and are tagged using the following syntax: **class_type=taxonomic**|{taxonomic base form}

247.
```
{base=Bacterium
    entry=E0456333
        cat=noun
        variants=inv
        varients=uncount
        class_type=taxonomic|Bacterium
```

}

15.5 Other variants

Beyond the specific class_types listed above, it is possible that in the future, other semantic groupings of terms will become useful or relevant. These groupings are currently subsumed by the **class_type=other** tag. Potential domains of interest include: genes, proteins, biological terms, clinical terms, medical terms, technical terms, or biomedical terms. When the **class_type=other** tag is used, the specific domain information will instead be given in the annotations slot.

Symbols

An index of example lexical entries.

A

B

C

D

O

P

Q

R

S

T