LEXICAL DATABASE ENRICHMENT THROUGH SEMI-AUTOMATED MORPHOLOGICAL ANALYSIS

Volume 2

THOMAS MARTIN RICHENS

Doctor of Philosophy

ASTON UNIVERSITY

January 2011

Contents

6 Evaluation 6

6.1 Measures of Semantic Relatedness for Word Sense Disambiguation 7

6.1.1 WordNet-based Measures 7

6.1.1.1 A Crude Measure 7

6.1.1.2 Direction Reversals 8

6.1.1.3 Taxonomic Depth 9

6.1.1.4 Extended Gloss Overlaps 11

6.1.1.5 Bag of Words 12

6.1.2 Evaluating WordNet-based Measures 13

6.2 Gold Standard Datasets 16

6.2.1 SENSEVAL 16

6.2.2 SENSEVAL-2 17

6.3 Adaptation of the Extended Gloss Overlaps Disambiguation Algorithm for Morphosemantic Wordnet Evaluation 18

6.3.1 Semantic Relatedness Measures 19

6.3.2 Relatives Lists 20

6.3.3 Gold Standard Data Set 21

6.3.4 Testbed 21

6.3.4.1 Disambiguator 22

6.3.4.2 Text Reader 22

6.3.4.3 Disambiguation Context Window 22

6.3.4.4 Window Occupants 24

6.3.5 Implementation of Semantic Relatedness Measures 24

6.3.6 Implementation of Disambiguation Algorithms 26

6.3.6.1 Generic Disambiguation Algorithm One by One 27

6.3.6.1.1 Target Disambiguation 28

6.3.6.1.2 Marking the Disambiguation Output 29

6.3.6.2 Differences between the One by One Generic Disambiguation Algorithm and Banerjee and Pedersen's Extended Gloss Overlaps 30

6.3.6.3 Nearest Neighbours Algorithm 33

6.3.6.4 Baseline Disambiguation by Frequency 33

6.4 Results 34

6.4.1 Execution Times 34

6.4.2 Performance Metrics 36

6.4.3 Performance 36

6.4.3.1 B&P Algorithm 37

6.4.3.2 Nearest Neighbours Algorithm 41

6.4.3.3 One by One Algorithm 41

6.4.3.4 One by One Algorithm with Fast Alternatives 46

6.4.4 Interpretation of Results 46

7 Conclusion 49
7.1 WordNet 50
7.2 Morphological Analysis and Enrichment 54
7.3 Evaluation 63
7.4 Future Research Directions 64
7.4.1 Applications of Derivational Morphology 66
References 69
URLs of Digital Resources 76
Class Diagrams 77
Class Diagram 1: Subclasses of Synset and WordSense 77
Class Diagram 2: Top Level Class Diagram of WordNet Model and Lexicon 78
Class Diagram 3: Revised Wordnet Design 79
Class Diagram 4: WordWrapper Structure 79
Class Diagram 5: Relations 80
Class Diagram 6: Lemmatiser 81
Class Diagram 7: Revised Lexicon Design 82
Class Diagram 8: Classes used to Represent CatVar Data and Morphological Rules 83
Class Diagram 9: Affix Tree 84
Class Diagram 10: Final Implementation of Affix Tree 85
Class Diagram 11: POSTaggedMorpheme 86
Class Diagram 12: WordBreaker 87
Class Diagram 13: Prefixation 87
Class Diagram 14: Disambiguator 88
Appendices 89
Appendix 1 Classes used to model WordNet and classes used in morphological analysis 89
Appendix 2 WordNet verb frames 100
Appendix 3 Ring topologies 101
Appendix 4 WordNet verb categories 102
Appendix 5 Valency and frame inheritance 102
Appendix 6 Valid inheritance by tightening selectional restrictions 105
Appendix 7 Evaluation of hypernym / troponym relations between verbal synsets in sample violating the relaxed rules for frame inheritance 106
Appendix 8 CatVar cluster members unrelated to headword 106
Appendix 9 Morphological rules formulated 108
Appendix 10 Original table of morphological rules (original version; §3) 121
Appendix 11 Words autogenerated from CatVar headwords but unrelated to them 126
Appendix 12 Productivity of morphological rules (CatVar dataset) 127
Appendix 13 Productivity of morphological rules (Word list dataset) 131
Appendix 14 Application of generalised spelling rules for suffix stripping 134

Appendix 15 Undergeneration in suffix stripping 136
Appendix 16 Candidate prefixes 139
Appendix 17 Candidate suffixes 141
Appendix 18 Properties of encoded lexical relations 144
Appendix 19 Formats of output files for morphological analysis 146
Appendix 20 Formats of input files for morphological analysis 149
Appendix 21 Suffixation Analysis Algorithm 150
Appendix 22 Relation types with their converses 151
Appendix 23 Preposition taxonomy by digraph analysis 152
Appendix 24 Preposition record fields 153
Appendix 25 Superordinate taxonomic categorizers 154
Appendix 26 Top ontology for prepositions 154
Appendix 27 Preposition antonyms 161
Appendix 28 Adjective to adjective pertainyms 162
Appendix 29 Exceptions specified in implementing the WordNet model 163
Appendix 30 Morphological rules for "-ion" suffix 163
Appendix 31 Morphological rules for "-al" suffix 164
Appendix 32 Morphological rules for "-ant" suffix 166
Appendix 33 Morphological rules for "-ent" suffix 167
Appendix 34 Morphological rules for "-ic" suffix 168
Appendix 35 Morphological rules for "-itis" suffix 168
Appendix 36 Complete morphological rules (final version; §5) 169
Appendix 37 Primary suffixation analysis results for "-able", "-ical" & "-ician" 181
Appendix 38 False lexical stems (Prefixation stem stoplist) 183
Appendix 39 Section from initial concatenation analysis results 186
Appendix 40 Concatenation first component stoplist 187
Appendix 41 Concatenation last component startlist 188
Appendix 42 Words starting with "non-" and "un-" which are not antonymous prefixations 192
Appendix 43 Antonymous prefixation exceptions and counter-exceptions 195
Appendix 44 1st. secondary suffix set as ordered by the optimal heuristic 197
Appendix 45 Homonyms with POS variation: result samples 198
Appendix 46 Secondary concatenation last component startlist 200
Appendix 47 Secondary concatenation complementary first component stoplist 200
Appendix 48 Secondary concatenation analysis results (complete) 201
Appendix 49 Irregular prefixes with sample instances 203
Appendix 50 Prefix translations 212
Appendix 51 1st. secondary prefix set as ordered by the optimal heuristic 224
Appendix 52 Linking vowel exceptions and reverse linking vowel exceptions 227
Appendix 53 Secondary suffix stripping stoplist 231
Appendix 54 Final suffixation reprieves 232
Appendix 55 Iterative suffixation analysis: input and output 236
Appendix 56 Iterative prefixation analysis: input and output 242
Appendix 57 Tertiary concatenation whole word stoplist 251
Appendix 58 Atomic dictionary 1/50 samples prior to stem processing 254
Appendix 59 Stem Dictionary Pruning Algorithm 256

Appendix 60 Stem meanings 258
Appendix 61 Encoding of relations between stems and their components 280
Appendix 62 Generic disambiguation Algorithm One by One 282
Appendix 63 Disambiguation results 286
Appendix 64 Mappings from *Claws* POS tags to the POSes of traditional grammar 290
Appendix 65 The WordNet model 291

Attached CD

Tables in Main Text

Table 52: Best SENSEVAL WSD scores compared to baseline 18
Table 53: Enumeration types specified by the disambiguator 26
Table 54: Configurations for consecutive disambiguation runs 27
Table 55: Sequential attempts at target disambiguation 28
Table 56: WSD execution times 35
Table 57: Distribution of relation types and lexical relations among relation type categories 56

Text Figures

Fig. 10: Disambiguation process diagram 23
Fig. 11: B&P WSD results 37
Fig. 12: WSD algorithms compared (window size 5) 39
Fig. 13: Nearest Neighbours WSD results 40
Fig. 14: WSD algorithms compared (window size 7) 42
Fig. 15: WSD algorithms compared (window size 11) 43
Fig. 16: One by One WSD results 44
Fig. 17: One by One WSD results with fast alternatives 45

6 Evaluation

The utility of the morphologically analysed lexicon would best be demonstrated by its deployment in an automatic translation application, either of the kind proposed by Habash (2002; §7.4.1) for Spanish to English translation, requiring more comprehensive resources at the target language end, or in conjunction with a second morphologically analysed lexicon for another language. As any such evaluation would clearly imply another research project, evaluation has focussed on the utility of the morphosemantic wordnet which combines the morphologically analysed lexicon with a preposition-enriched version of WordNet, at a task for which WordNet has widely been deployed and which is a requirement for most more complex NLP applications, namely word sense disambiguation (*WSD*).

The next section reviews various approaches to WSD. The approaches discussed all select senses of words based on their relatedness or similarity to other words in a context[1]. A measure is therefore needed of the relatedness or similarity of any pair of concepts. Various measures are discussed before the Extended Gloss Overlaps approach (Banerjee & Pedersen, 2002; 2003; §6.1.1.4) is adopted. Evaluation of performance at WSD requires a *gold standard dataset*. Two SENESVAL datasets are discussed in §6.2 of which SENSEVAL-2 is adopted. §6.3 describes the implementation of an adaptation of the Extended Gloss Overlaps Disambiguation Algorithm for the evaluation of the morphosemantic wordnet, such that the contribution to WSD of WordNet relations and lexical relations based on derivational morphology can be compared. Because of the greediness of the algorithm as described by Banerjee & Pedersen (2002; 2003), some variants upon it are also presented. In line with Kilgarriff's (1998a; 1998b; §6.2) recommendations, disambiguation by corpus frequency is also implemented as a baseline for the evaluation. The results of the evaluation with all the variant algorithms are presented in §6.4.

[1] For the distinction between relatedness and similarity, see §6.1.2.

6.1 Measures of Semantic Relatedness for WSD

Lesk (1986) came up with a proposal to disambiguate words by comparing their glosses in a machine-readable dictionary with those of other words in a context window and counting the common words (measuring the gloss overlap). That sense of any word whose gloss has the greatest overlap with those of its neighbours in the context window is then the sense chosen. The quality, and in particular the comprehensiveness, of the dictionary used clearly will have an impact on the results. Lesk reports an accuracy of 50-70%, using the Oxford Advanced Learner's dictionary, applied to examples from *Pride and Prejudice* and an *Associated Press* news story, using a window size of 10 words. Lesk goes into little detail about the methodology and reaches no conclusion on the optimum window size or, once a word has been disambiguated, whether only the gloss for the sense discovered should then be used for disambiguating other words (§6.3.6.1.1). This algorithm has been extended by Banerjee & Pedersen (2002; 2003; §6.1.1.4) and further extended for the evaluation of the morphosemantic wordnet (§6.3).

6.1.1 WordNet-based Relatedness Measures

6.1.1.1 A Crude Measure

The simplest possible WordNet-based similarity measure counts the shortest distance between the nodes representing the synsets to which the word senses being compared belong. This crude measure can be written mathematically as:

$$rel(c_1, c_2) = -len(c_1, c_2)$$

where c_1 and c_2 are 2 concepts (synsets).

There are two main problems with this measure:

1. The path traversed through WordNet between synsets may include links in opposite directions: this is addressed by Hirst & St-Onge (1998; §6.1.1.2).

2. Not all links between WordNet synsets represent the same semantic distance: this is addressed by Stetina & Nagao (1997) and Leacock & Chodorow (1998; §6.1.1.3) by introducing the concept of *taxonomic depth*.

An attempt at using the crude measure for disambiguation within the current research project was abandoned because of the long execution time required.

6.1.1.2 Direction Reversals

Hirst & St-Onge (1998) introduce the idea of *lexical chains*, based on WordNet, which they apply to the detection of malapropisms. A lexical chain is a sequence of words from a context (not necessarily in the same order in which they occur in the context), the links between which are weighted. The idea is that a lexical chain links words taken from a context with links weighted by strength. The following levels of strength are recognised:

- Very strong: the same word;
- Strong: linked by an ANTONYM, SIMILAR or SEE_ALSO relation;
- Medium-strong: linked by an allowable path through WordNet viewed as a graph;
- Weak: linked, but not by an *allowable path*, and having a weighting of zero.

The concept of an allowable path depends on conceiving of a wordnet as a set of interconnected upside-down trees, where *upward* means towards the root, and *downward* means towards the leaves. A *horizontal* link is a link between trees, or between branches of the same tree. An allowable path is defined as a path comprising between 2 and 5 links between synsets defined by the following rules:

- no other direction may precede an upward link;
- at most one change of direction is allowed except where a horizontal link occurs between an upward and a downward direction.

A medium-strong relation is weighted by the following equation:

$$w = C - l - kd$$

where w is the weight, l is the length of the path, d is the number of direction changes and C and k are constants. Weak links are rejected for lexical chaining. The

weighting of a medium-strong relation is a semantic relatedness measure. Unfortunately, the weightings of the very strong and strong categories are not given in their paper, nor are values for C and k, though Budanitsky & Hirst (2006; §6.1.2) used values $C = 8$ and $k = 1$. The concept of direction reversals is applicable to morphological relations between words as encoded in the morphosemantic wordnet though not to directionless WordNet relations, including the original WordNet DERIV relation, to which this measure cannot be applied. If very strong links always override the others and strong links always override medium-strong, then this relatedness measure could be applied to the morphosemantic wordnet, and the value of C could be varied according to an assessment of the importance of each relation type.

6.1.1.3 Taxonomic Depth

Stetina & Nagao (1997) propose a WordNet-based measure of semantic distance

$$D = \frac{\left(\frac{L_1}{D_1} + \frac{L_2}{D_2}\right)}{2}$$

where L_1 and L_2 are the lengths of the paths from 2 synsets to their nearest common ancestor, and D_1 and D_2 are the distances of the same 2 synsets from the root of the taxonomy.

Leacock & Chodorow (1998) propose another WordNet-based similarity measure

$$sim_{ab} = \max\left[-\log\left(\frac{N_p}{2D}\right)\right]$$

where N_p is the number of synsets on the path from a to b and D is the maximum depth of the taxonomy.

The concept of depth in both these equations presupposes positing a root node as the HYPERNYM of all the unique beginners of each POS taxonomy, which should ensure that there is a path between every synset of the same POS, except for modifiers, as well as a path from each synset to the root node, which allows depth to

be calculated. In practice this does not work for all synsets because of some anomalies of WordNet as follows:

1. Modifiers in WordNet do not participate in HYPERNYM/HYPONYM relations (This does not apply to the pruned model of WordNet developed as precursor to the morphosemantic wordnet where the SIMILAR relation type between adjectives has been replaced; §4.3.2).
2. There are nouns (especially proper nouns) in WordNet which do not participate in HYPERNYM/HYPONYM relations, but are free-floating, connected only by INSTANCE relations (§2.2.2.2.6). This has also been corrected in the pruned model of WordNet but only where there can be certainty that a noun is a proper noun (§4.3.4).
3. There is no common root for the WordNet verb taxonomy (§2.2.2.2.6).

In practice, Leacock & Chodorow (1998) and Budanitsky & Hirst (2006) only apply this measure to nouns.

The depth variable is meaningless with reference to lexical relations between words unless we posit a similar root node which connects every word root, many of which are not represented by any Synset but only as stems in the stem dictionary (§5.3.10). Hence this measure is unsuitable for application to the evaluation of the morphosemantic wordnet.

All these WordNet-based measures are refinements of the crude one and share the same problem: if the word senses being compared do not share the same word POS, there will most likely be no shortest path between the two. This means that strongly related words from different classes would have a calculated semantic distance of infinity. In the morphosemantic wordnet, there are many links across POS boundaries and the measure could better be applied, but the comparison with the non-morphologically-enriched version would be almost meaningless.

6.1.1.4 Extended Gloss Overlaps

Banerjee & Pedersen (2002) extend the approach of Lesk (1986), applying it using the glosses in WordNet, but instead of taking into consideration only the glosses of the senses of the words in the context window, they also take into account the glosses of their WordNet relatives. They also modify the scoring mechanism by assigning greater weights to overlapping sequences of more than one word, such that the weight of the overlap is equal to the square of the number of words in the overlap. Overlaps consisting entirely of "non-content words" (undefined) are ignored. They use a small window, whose size is an odd number, in which the target (the word to be disambiguated) is in the middle, except at the beginning or end of the available context, where they use an asymmetrical window of the same size. They evaluate every possible combination of a sense of the target word, or sense related to a target sense by a WordNet relation, with the senses, or similarly related senses, of the other words in the window, by summing the gloss overlap scores of each pair within each combination. They then select the sense of the target word which occurs in the highest scoring combination. The best senses of the other words are discarded. The identified sense of the target is not recycled for use in subsequent disambiguations[2]. The WordNet relations used are HYPERNYM, HYPONYM, HOLONYM, MERONYM and ATTRIBUTE. The senses of a word examined are limited to those of the POS of the word, where this is provided. Where two senses of the target word achieve an equal score, the one which has the greatest frequency is chosen by default. An overall accuracy of 31.7% is reported from tests applied to 73 target words within 4328 instances, taken from SENSEVAL-2. This compares with 12% if POS-tags are ignored or 16% from applying another variant of the Lesk Algorithm (without WordNet relations) to the same data.

Banerjee & Pedersen (2003) extend their experiments to use more WordNet relation types including SIMILAR and SEE_ALSO. To reduce noise, function words, defined as pronouns, prepositions, articles and conjunctions, are now excluded from the beginning and end of the gloss overlaps. Function words are also removed from the contexts, prior to defining a window of size 3. In cases where there is more than one

[2] This issue is taken up in §6.3.6.1.1.

equally good best sense for a target word, frequency is no longer used as a tie breaker but all best senses are reported and partial credit is given. In a fresh evaluation, precision is defined as the number of correct answers divided by the number of answers and recall is defined as the number of correct answers divided by the number of test cases. A precision of 35.1% and a recall of 34.2% are now reported against a baseline which selects word senses randomly, which gives precision and recall of 14.1%. These results are superior to two out of the three best performing fully automatic unsupervised systems which participated in the original SENSEVAL-2 contest (§6.2.2). Banerjee & Pedersen report that increasing the window size to 5, 7, 9 or 11 does not significantly improve the results. They also report that using limited subsets of WordNet relation types results in significant deterioration in performance.

An extension and adaptation of Banerjee & Pedersen's algorithm to the evaluation of the morphosemantic wordnet is presented in §6.3.

6.1.1.5 Bag of Words

Sinha et al. (2006) propose an innovative similarity measure for WSD which uses a wide window comprising the sentence containing the word *w* to be disambiguated plus the preceding and following sentences, all the words in which comprise a bag of words set *C*. For each sense *s*, of *w*, a second bag of words set *B* is created comprising:

- the synonyms of *s*;
- the glosses for the synset *S* comprising *s* and its synonyms;
- the usage examples for *S*;
- the words in the synsets which are relatives of *S* by a direct or indirect HYPERNYM, HYPONYM OR MERONYM relation from *S*;
- the glosses for those relatives;
- the usage examples for the relatives;

The size of the intersection of sets *B* and *C* is measured, and the sense *s* for which the corresponding set *B* has the greatest intersection with *C* is the sense assigned to *w*.

This measure could be adapted for application to the morphosemantic wordnet by using the above measure as a control, with a purely morphological measure for comparison comprising:

- the words in the synsets which contain direct or indirect morphological relatives of the words in *S*;
- the glosses for those synsets;
- the usage examples for those synsets,

and a morphosemantic measure combining the morphological measure with that of Sinha et al., 2006.

6.1.2 Evaluating WordNet-based Measures

Budanitsky & Hirst (2006) review a number of WordNet-based measures of semantic relatedness and apply tests to determine which are best. They make a distinction between *relatedness* and *similarity*. These measures can be represented as two different scales on which, for both, synonymy has a value of 1, but antonymy has a value of 0 on the similarity scale but a value of 1 on the relatedness scale, where 0 represents completely unrelated. However, when making their comparisons, they do not attempt to convert 1 measure to the other. They consider Hirst & St-Onge's (1998) measure to be a relatedness measure, while all the others they discuss are similarity measures.

Two types of tests are proposed: the first is based on comparisons with human ratings of the relatedness of word pairs and the second on the ability to detect and correct malapropisms. Because of the cost of obtaining human ratings, the authors rely on two existing studies (about which they give few details) and compare these with the results for the same sets of word pairs obtained from the measures being tested, which in several cases means simply re-reporting the results given by their authors. The comparisons with the two different existing studies give widely disparate results. Budanitsky & Hirst acknowledge many shortcomings of these tests, particularly the small size of the datasets and the fact that the human subjects were given words to assess rather than word senses.

The test on malapropisms was twofold. The measures being compared were applied first to identifying malapropisms from the lack of relatedness of words in a context, and then to finding a word more related to the context which could be seen to be its correction. The malapropisms were deliberately introduced into the test text, so that the right correction was always known. This methodology was originally proposed by Hirst & St-Onge (1998), whose relatedness measure is one of the contestants.

Although Budanitsky & Hirst describe some non-WordNet-based measures, all the measures tested are WordNet-based. These fall into two main categories, those which use only data found in WordNet, and those which also use a sense-tagged corpus. While the corpus-based approaches are of interest, they have not been considered as possibilities for testing the morphosemantic wordnet, because of the time taken by such experiments, given the time available for the evaluation and the paucity of corpora tagged with WordNet 3.0 senses.

Of those measures which use only WordNet data, only two are evaluated. It is unfortunate that the crude measure is not evaluated, as it would provide an informative baseline. However all the other measures are refinements of the crude one. In practice, though it is not specifically stated, it appears that Budanitsky & Hirst only looked at nouns. This is explicit for the human ratings as all the test word pairs are given.

Budanitsky & Hirst discuss the variables used by the various measures, including direction reversals (§6.1.1.2) and taxonomic depth (§6.1.1.3). Another variable is the lowest superordinate of 2 synsets (most specific common subsumer), whose applicability again depends on the directionality of the relations, though it is unclear how this should be determined where there is a combination of HOLONYM/MERONYM relations and HYPERNYM/HYPONYM relations. In practice, it appears, though it is not explicitly stated, that most of the measures only use HYPERNYM/HYPONYM relations, except for the direction reversals measure, which also uses HOLONYM/MERONYM relations.

The inapplicability of some of the variables means that the measures which use them cannot be applied to the morphosemantic wordnet. The crude measure and direction

reversals are clearly applicable. The remainder all require a depth variable. Although this could be computed, it is not sufficiently meaningful in the context of lexical relations to be worth pursuing. Of the two applicable measures, only Hirst and St. Onge's direction reversals measure is evaluated. On one of the two tests based on human ratings, the direction reversals measure gives the poorest performance of all 5 measures evaluated and on the other it outperforms 2 out of 3 sense-tagged corpus-based measures, but is beaten by the other and by another measure which uses the depth variable but not the lowest superordinate variable; for malapropism detection it gives the poorest recall but good precision, being clearly beaten by only one corpus-based measure; for malapropism correction it again gives the poorest recall and precision is disappointing as it beats only one corpus-based measure. Hirst and St. Onge's direction reversals measure assigns a relatedness value of 0 to pairs which fail to satisfy the criteria for an allowable path. Budanitsky & Hirst believe that without this cutoff, it might have performed better at the human ratings evaluations, especially as it is the only measure discussed which makes use of HOLONOM/MERONYM relations and the only one designed to test relatedness rather than similarity.

Since Hirst and St. Onge's direction reversals measure is the only applicable one evaluated, the choice of measure for evaluating the morphosemantic wordnet cannot take the results of Budanitsky & Hirst's evaluation into account. The other applicable measures are the crude measure (which has been experimented with, but proved very slow to execute) and that of Sinha et al. (2006), but the final choice was to adapt Banerjee & Pedersen's (2002; 2003) measure. The main consideration here, apart from the meaningfulness of variables in the context of a morphologically enriched WordNet, was the need to run tests in the time available. An implementation of Hirst and St. Onge's measure would be an interesting area for future research, and might well turn out to be faster than the crude measure, as it would not be necessary to navigate paths through the network which do not conform to the directionality rules. The method described by Sinha et al. (2006; §6.1.1.5) would also be an interesting area to investigate.

6.2 Gold Standard Datasets

Kilgarriff[3] (1998a, 1998b) discusses the pitfalls of developing gold standard datasets for evaluating WSD programs. He raises the issue of upper and lower bounds to the possible performance of a WSD System. The upper bound is largely determined by the validity of the sense distinctions and the consistency of the semantic relations; the lower bound (*baseline*) is the performance of a naive system which always selects the sense with the highest recorded corpus frequency. This appropriate baseline is ignored in the evaluation of their own work by Banerjee and Pedersen (2002; 2003; §6.1.1.4), even though they use it as a tie breaker. This baseline is however compared with results obtained both by reproducing and by extending their methodology in the evaluation of the morphosemantic wordnet (§6.4).

6.2.1 SENSEVAL

Kilgarriff also cites the contribution of Resnik & Yarowsky (1997), whose proposals were largely incorporated into the development of the original *SENSEVAL* dataset. One proposal was that WSD should not be evaluated as simply right or wrong, but there should be gradations of how near the WSD output is to the gold standard. In the discussions which ensued at the SIGLEX workshop, there emerged a difference of opinion between computer scientists, who wanted a fixed set of dictionary definitions to work with, and lexicographers, whose main concern was getting inter-annotator agreement, over the particular issue of whether to allow multiple taggings for a single word. The conclusion was that multiple taggings should be allowed but only as a last resort.

In order to maximise inter-annotator agreement, lexicographers were employed, rather than volunteers, and they were allowed to confer when they disagreed, in order to arrive at a consensus. The quest for an internally consistent set of word senses disfavoured WordNet and favoured the *HECTOR* dictionary, based on the 20-million word BNC pilot corpus. Mappings were provided from HECTOR senses to WordNet

[3] despite his disbelief in word senses (§2.1.1).

senses for systems which only have access to the WordNet senses. The most accurate and consistent sense-tagging is achieved when it concentrates on words with a large number of instances in the text, which are likely to illustrate different meaning, rather than tagging a large number of unrelated words. It is also better when the taggers work one word at a time so that they are looking at the same set of definitions, rather than proceeding sequentially through the text. These are reasons for tagging relatively few selected words in the text and using these for WSD evaluation.

6.2.2 SENSEVAL-2

For SENSEVAL-2, WordNet was chosen as the English lexicon, disregarding the reasons for which it was rejected for SENSEVAL-1 (§6.2.1). Edmonds & Cotton (2001) state that 90% inter-annotator agreement was the goal, but say nothing about how far this goal was achieved. The taggers were volunteers. These facts raise doubts about SENSEVAL-2 as a gold standard. There were two WSD tasks: a lexical samples task and an all words task. Multiple taggings were allowed, and gradations of results between right and wrong. These gradations are not mentioned by Banerjee and Pedersen (2002; 2003; §6.1.1.4) nor are they reflected in the SEMCOR format version used for evaluating the morphosemantic wordnet (§6.3.3). Measures of recall and precision were defined: recall as percentage of right answers out of all instances in the test set and precision as percentage of right answers out of all answers given. Coverage was defined as the percentages of answers given out of all instances (§6.4.2).

Edmonds & Kilgarriff (2002) report the best scores for the SENSEVAL-1 and SENSEVAL-2 evaluation exercises, against a baseline of selecting the most frequent sense in an unspecified corpus (Table 52; §§6.3.6.4, 6.4.3, 6.4.4). It is notable here that the best score is lower on SENSEVAL-2. Edmonds & Kilgarriff say that this has been variously attributed to the use of WordNet senses or to a dataset which was more difficult to disambiguate. It is unclear why the SENSEVAL-2 baseline is lower for unsupervised systems.

Table 52: Best SENSEVAL WSD scores compared to baseline

Dataset		Systems	Baseline	Best Score
SENSEVAL-1	Lexical sample		57%	78%
SENSEVAL-2	Lexical sample	Supervised	48%	64%
		Unsupervised	16%	40%
	All words		57%	69%

6.3 Adaptation of the Extended Gloss Overlaps Disambiguation Algorithm for Morphosemantic Wordnet Evaluation

The main objective of this evaluation is not to find the best disambiguation algorithm, though this question is elucidated as a by-product of the tests (§6.4.4), nor to make a judgement about WordNet senses distinctions (§2.1), though the results inevitably also reflect on this. The main objective is simply to establish whether the morphologically enriched version can outperform WordNet at a WSD task.

A WSD algorithm based on a measure of semantic relatedness between pairs of word senses has been described by Banerjee & Pedersen (2002; 2003; §6.1.1.4). This algorithm is here adapted to use additional new measures of semantic relatedness (§§6.3.1, 6.3.5).

One shortcoming of Banerjee & Pedersen's algorithm has been noted (§6.1.1.4), namely its failure to recycle the identified sense of the target word when disambiguating the other words, so that the identified sense of a second target word within the same window may be inconsistent with that of the first. Mutual disambiguation of the words in a moving window would be likely to give more consistent results but would be more demanding programmatically and in terms of computational resources. Moreover the results would be less comparable with those of Banerjee & Pedersen. Mutual disambiguation will not be implemented in this exercise, but the sense inconsistencies will be recorded as *paradoxes* (§6.3.6.1.1).

Window size is an important variable: Lesk (1986; §6.1) favours larger windows; Banerjee & Pedersen favour smaller windows. Experiments will be described with a variety of window sizes (§6.4).

6.3.1 Semantic Relatedness Measures

The proposed measures of semantic relatedness of two word senses are all new except for the last which is that used by Banerjee & Pedersen:

1. The first measure gives a score of 2 if both word senses are included in each other's relatives' lists (§6.3.2), or 1 if only one of the words is included in the other's relatives' list, otherwise 0.

2. The second measure gives a score equal to the number of common members of the 2 relatives' lists.

3. The third measure calculates the gloss overlaps, as described by Banerjee & Pedersen (§6.1.1.4) between each word sense and each relative in the other's relatives' list, and gives a score equal to the sum of the gloss overlaps.

4. The fourth measure calculates the gloss overlaps between each relative in one relatives' list and each relative in the other relatives' list, and gives a score equal to the sum of the gloss overlaps[4].

These measures compare the relatives lists of a sense of the target with those of another window occupant. Measures 1-3 are *fast alternatives* to Banerjee & Pedersen's measure. Of these measures, the first is the strongest indicator of semantic relatedness, but the least likely to give a score > 0. At no point is the score from any of these measures to be compared with the score from any other as they are non-comparable. The same measure is to be applied for every word sense comparison between senses of the target word and senses of other words in the window. If a single comparison returns a maximum score, then the sense of the target involved in that

[4] as in Banerjee & Pedersen's work.

comparison will be selected as its best sense. If the measure returns a score of 0 for every comparison, or if more than one comparison returns the same maximum score with that measure, then the target cannot be disambiguated using that measure. Only when the target cannot be disambiguated using one measure will the next measure is adopted. The measures are to be applied successively to each target disambiguation operation, until the application of one of them can establish a best sense for the target (§6.3.6.1.1).

6.3.2 Relatives Lists

The main objective is to compare the effect of applying the same semantic relatedness measures using WordNet relations only, lexical relations only and both in combination. This requires the compilation of lists of semantic and morphological relatives. A `RelativesList` specifies a set of relations for a `WordSense` and a set of synsets implied by those relations. There are two subtypes.

- A `SemanticRelativesList` encapsulates a relations set which combines the `Set<Relation>` of the specified `WordSense` along with the `Set<Relation>` of the `Synset` which contains it. Its set of synsets is the set of the targets of the relations set (§1.3.2).
- A `LexicalRelativesList` specifies a set of lexical relations (§3.5.3) and has three subtypes:
    - a `DirectLexicalRelativesList` is never used because the set of direct lexical relations for any sense of a given word will always be the same and so will not be an aid to WSD;
    - a `SynonymLexicalRelativesList` encapsulates a relations set which combines the `Set<Relation>` of the `GeneralLexicalRecord` and the `Set<Relation>` of the `POSSpecificLexicalRecord` of every word in the `Synset` which contains the specified `WordSense`;
    - a `SemanticRelativesLexicalRelativesList` encapsulates a relations set which includes all the relations in a `SynonymLexicalRelativesList` plus the `Set<Relation>` of the `GeneralLexicalRecord` and the `Set<Relation>` of the

`POSSpecificLexicalRecord` of every word in every `Synset` in the `SemanticRelativesList` for the `WordSense`.

The set of synsets of a `LexicalRelativesList` comprises every `Synset`, which is mapped to by a `LexicalRecord` (§3.5.3) corresponding to the target of any of the relations.

6.3.3 Gold Standard Data Set

Unfortunately the mappings available from HECTOR senses to WordNet senses do not apply to WordNet 3.0, whose senses are used in the morphosemantic wordnet and so the original SENSEVAL dataset (§6.2.1) could not be used for its evaluation. Instead the SEMCOR format versions of SENSEVAL-2 all words task with WordNet 3.0 senses (http://www.cse.unt.edu/~rada/downloads.html) was chosen as the best available compatible alternative, despite the evidence suggesting that the high standards applied in devising the original SENSEVAL exercise have been largely disregarded (§6.2.2).

Banerjee and Pedersen used SENSEVAL-2 for their evaluation (§6.1.1.4) and so it seemed that it would be possible to make a comparison with their findings. It emerged, only after selecting the dataset, that Banerjee and Pedersen used the lexical samples task and not the all words task for their evaluation (§6.2.2). This dataset was not available in the same format, but it is still of interest to compare their findings with results using their method, applied to the all words task.

6.3.4 Testbed

For the relationships between classes which are used to implement the disambiguator, please refer to Class Diagram 14.

6.3.4.1 Disambiguator

The `Disambiguator` has two main components as follows:

- `GoldStandardReader reader;`
- `DisambiguationContextWindow window;`

6.3.4.2 Text Reader

A `GoldStandardReader` handles the test dataset, passing on as much information to the `DisambiguationContextWindow` as is allowed for the test being conducted (Fig. 10). This will always include the text content and which words are to be disambiguated, but may or may not include other information, in particular the POS of each word and its lemma, depending on the specification of the test. The correct senses of the words are never passed to the `DisambiguationContextWindow`. Each time the window is advanced, a `DisambiguationOutputWord` encapsulating the word leaving the window and its disambiguated sense is stored, eventually to be passed back to the `DisambiguationTextReader` for marking (§6.3.6.1). The `GoldStandardReader` encapsulates a buffer with file input facilities along with a list of stop words[5] which are not allowed to pass through to the `DisambiguationContextWindow`. To minimise noise from irrelevant senses, prepositions are allowed only if they are specified as disambiguable.

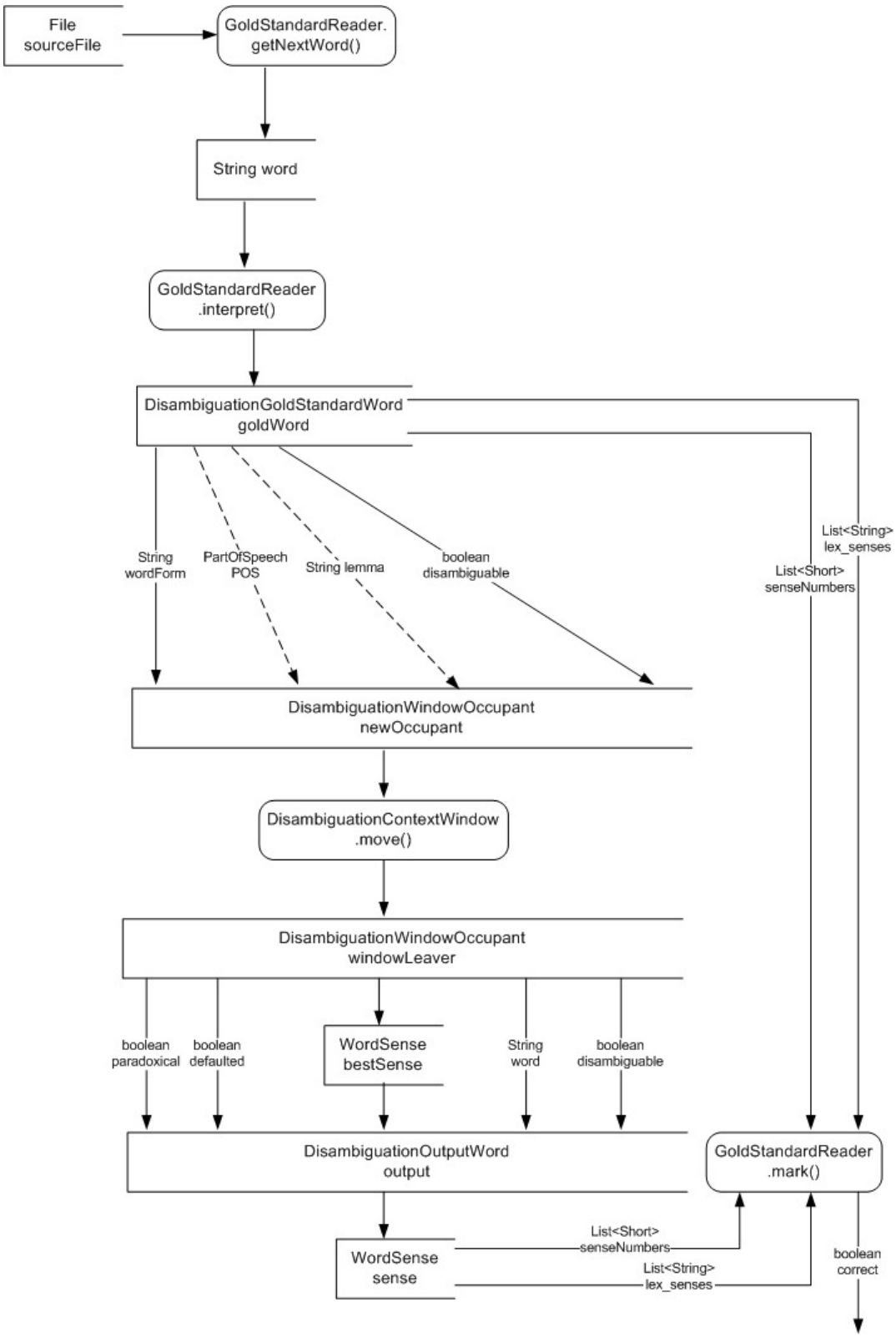
6.3.4.3 Disambiguation Context Window

The `size` field of the single `DisambiguationContextWindow` must be defined at the outset and remain constant thereafter. The window size must be an odd number otherwise the target will not be at the centre of the window.[6] Fields `morphologicalAwareness`, `currentLexicalRelativity`, `senseMatchMeasure`

[5] "am", "is", "are", "was", "were", "being", "been", "has", "had", "having", "no", "any", "some", "every", "more", "most", "very", "too", "rather", "the", "a", "an", "this", "that", "these", "those", "it", "'s", "'d", "can", "will", "shall", "'ll".

[6] The window occupants are represented as a `LinkedList<DisambiguationWindowOccupant>`, which remains constant in size except between the addition and removal of an occupant, which are consecutive operations. The target position in the window is identified by an index set to `size` / 2 (by integer division), except in experiments where the target position varies at the beginning and end of the

Fig. 10: Disambiguation process diagram



text (§6.3.6.2). As the target index remains constant, the performance of these consecutive operations has the effect of moving each occupant along by one place in the window so that each occupant in turn is the target at the mid-point of its lifecycle.

and `glossOverlapMeasure` must be defined at instantiation, but can be changed so that the same window can be re-used on the same text with different settings. These fields are instances of enumeration types `MorphologicalAwareness` and `LexicalRelativity` (Table 53) and classes `SenseMatchMeasure` and `GlossOverlapMeasure` respectively, both of which are subclasses of `SemanticRelatednessMeasure` (§6.3.5).

6.3.4.4 Window Occupants

A `DisambiguationWindowOccupant` represents a word within the window. When a new occupant enters the window, the next word must be provided by the `GoldStandardReader`, which must also specify whether the word is to be disambiguated. The lemma and POS may or may not be specified. If they are specified, they are assigned to fields `bestLemma` and `bestPOS`. If the POS is not specified, then field `possiblePOSes` is populated with all the POSes found in the lexicon for the word. If the lemma is not specified, then field `possibleLemmas` is populated with the lemmas returned by the `Lemmatiser` and field `possibleSenses` is populated with every `WordSense` returned by the `Lexicon` for every lemma. If the lemma is specified then `possibleSenses` is populated with every `WordSense` returned by the `Lexicon` for the lemma (as the specified POS if any).

6.3.5 Implementation of Semantic Relatedness Measures

`SenseMatchMeasure` and `GlossOverlapMeasure` are subclasses of `SemanticRelatednessMeasure`, which specifies a *light* method[7] and a *heavy* method[8] (Table 55).

The light method returns a relatedness score obtained by comparing parameter `thisSynset` to each member of a `Collection<Synset> otherSynsets` added to a relatedness score obtained by comparing `otherSynset` to each member of

[7] `float measure(Synset thisSynset, Synset otherSynset, Collection<Synset> theseSynsets, Collection<Synset> otherSynsets)`
[8] `float measure(Collection<Synset> theseSynsets, Collection<Synset> otherSynsets)`

`theseSynsets`. The heavy method returns a relatedness score obtained by comparing each member of one `Collection<Synset>` to each member of another.

These two methods are implemented differently by a `SemanticRelatednessMeasure` and a `GlossOverlapMeasure` so that four methods implement the measures listed in §6.3.1.

`GlossOverlapMeasure` corresponds to the original Lesk (1986) Algorithm (§6.1); refinements have been implemented and tested in the following subclasses:

- `PhraseAwareGlossOverlapMeasure` extends `GlossOverlapMeasure`, implementing Banerjee & Pedersen's (2002; §6.1.1.4) variant on the basic algorithm such that the gloss overlap between any pair of glosses is not simply the number of words in common, but the weighted sum of the squares of the number of words in each overlap;

- `LengthAndPhraseAwareGlossOverlapMeasure` extends `PhraseAwareGlossOverlapMeasure`, implementing the suggestion, that the likelihood of a gloss overlap increases with the length of the glosses. The gloss overlap is that calculated by a `PhraseAwareGlossOverlapMeasure` divided by the average number of words in the two glosses;

- `SizeAndLengthAndPhraseAwareGlossOverlapMeasure` extends `LengthAndPhraseAwareGlossOverlapMeasure` and develops the same idea further by also taking into consideration the fact that the more glosses there are, the more likely a gloss overlap is to occur. The gloss overlap is that calculated by a `LengthAndPhraseAwareGlossOverlapMeasure`, but the `measure` methods return the summed overlaps divided by the average size of the two synset collections.

During preliminary testing on random scraps of text, it was found that classes `LengthAndPhraseAwareGlossOverlapMeasure` and `SizeAndLengthAndPhraseAwareGlossOverlapMeasure` did not perform any better

than PhraseAwareGlossOverlapMeasure while PhraseAwareGlossOverlapMeasure performed better than the base class GlossOverlapMeasure. Consequently all subsequent tests were performed using a PhraseAwareGlossOverlapMeasure.

6.3.6 Implementation of Disambiguation Algorithms

The concepts listed in the first column of Table 53 are essential to the comparisons made during the evaluation. *Lexical Relativity* specifies the kind of LexicalRelativesList to be used, if any (§6.3.2); *Morphological Awareness* specifies whether a SemanticRelativesList or a LexicalRelativesList is to be used[9]; the various *disambiguation algorithms* are described in §6.3.6.

Table 53: Enumeration types specified by the disambiguator

Lexical Relativity (table 55)	NON_LEXICAL	SYNONYMOUS	SEMANTICALLY RELATED	
Morphological Awareness (§6.3.6.1.1)	SEMANTIC	LEXICAL	MORPHO-SEMANTIC	
Disambiguation algorithm (§6.3.6)	ONE BY ONE	NEAREST NEIGHBOURS	B AND P	BASELINE

Prior to running any disambiguation experiment:

- The GoldStandardReader must input the marked-up text and identify its component words.
- The Disambiguator and its DisambiguationContextWindow must be instantiated, specifying the size of the window and whether or not it is allowed to know the lemmas and POSes of the words to be disambiguated.
- A suitable data structure must be set up to house the output, at its most simple, a List<DisambiguationOutputWord>.
- The window's currentLexicalRelativity and morphologicalAwareness fields must be defined. In practice, for most experiments, 5 consecutive disambiguation runs were performed with the configurations listed in Table

[9] In this context, SEMANTIC means that a SemanticRelativesList is to be used; LEXICAL means that a LexicalRelativesList is to be used and MORPHO-SEMANTIC means that both are to be used.

54. By varying the parameters, the same *generic disambiguation algorithm* can be applied to disambiguate the same text with each of these 5 configurations.

Table 54: Configurations for consecutive disambiguation runs

Position in Sequence	**Morphological Awareness**	**Lexical Relativity**	**Relations used**
1	SEMANTIC	NON LEXICAL	Wordnet relations only
2	LEXICAL	SYNONYMOUS	Lexical relations of synonyms
3	LEXICAL	SEMANTICALLY RELATED	Lexical relations of Wordnet relatives
4	MORPHO-SEMANTIC	SYNONYMOUS	Wordnet relations and lexical relations of synonyms
5	MORPHO-SEMANTIC	SEMANTICALLY RELATED	Wordnet relations and lexical relations of Wordnet relatives

6.3.6.1 Generic Disambiguation Algorithm One by One

In its simplest and original form, the generic disambiguation algorithm (pseudocode in Appendix 62) populates the window with occupants created by the `GoldStandardReader` with the permitted fields (§6.3.4.2) of the first words in the text. The procedure for advancing the window comprises four operations:

- A new `DisambiguationWindowOccupant` enters the window as if from the right.
- The oldest `DisambiguationWindowOccupant` leaves the window as if to the left.
- The `DisambiguationWindowOccupant` in target position[10] is disambiguated with reference to the other window occupants (§6.3.6.1.1).
- A `DisambiguationOutputWord` is created from the `DisambiguationWindowOccupant` leaving the window and stored in the output until the whole text has been disambiguated, when it is passed back to the `DisambiguationTextReader` for marking (§6.3.6.1.2).

This procedure is repeated until the text from which the `GoldStandardReader` supplies the words to window occupants is exhausted. Thereafter null window

[10] once the first `DisambiguationWindowOccupant` has reached the target position.

occupants enter the window until all the valid window occupants have left the window. Disambiguation ceases when the first `null` enters the target position.

6.3.6.1.1 Target Disambiguation

Each time the window is advanced, up to 4 consecutive attempts are made to disambiguate the target (Table 55). The algorithm proceeds to the next attempt only if the previous attempt has returned a null result.

Table 55: Sequential attempts at target disambiguation

Attempt	`Relatedness Measure`	Weight (§6.3.5)	Method
1	`Sense Match Measure`	Light	`measure(thisSynset, otherSynset, theseSynsets, otherSynsets)`
2	`Sense Match Measure`	Heavy	`measure(theseSynsets, otherSynsets)`
3	`Phrase Aware Gloss Overlap Measure`	Light	`measure(thisSynset, otherSynset, theseSynsets, otherSynsets)`
4	`Phrase Aware Gloss Overlap Measure`	Heavy	`measure(theseSynsets, otherSynsets)`

The idea behind the 4 attempts to disambiguate is to use, if possible, the faster `senseMatchMeasure`, which is a stronger indicator of semantic relatedness, only resorting to a `glossOverlapMeasure` in the absence of a sense match (§6.3.1). A light method requiring fewer synset comparisons is preferred where a result can be obtained from it.

At each attempt, the target is provisionally disambiguated with reference to each other `DisambiguationWindowOccupant` in turn. This provisional disambiguation is performed by comparing every possible `WordSense` of the target with every possible `WordSense` of the other `DisambiguationWindowOccupant`. That pair of senses is selected which attains the highest score from applying the specified `measure` method of the specified `SemanticRelatednessMeasure` (Table 55) using the `RelativesList` for each sense. The type of `RelativesList` is determined by the value of the `morphologicalAwareness` field (Table 54): if

MorphologicalAwareness is LEXICAL, then the LexicalRelativesList is used; if MorphologicalAwareness is SEMANTIC, then the SemanticRelativesList is used; if MorphologicalAwareness is MORPHO_SEMANTIC then both are used. Whichever measure method is being used (§6.3.5), each synset collection required as a parameter is provided by the corresponding RelativesList. If a light method is being used, the individual synsets required are those which contain the two senses being compared. If, at the fourth attempt, still no result is obtained (all the lists generated were null), then the default baseline disambiguation by frequency is executed and the occurrence of a default is recorded.

The selected sense of the target is assigned to the bestSense field of the target.[11] The other selected sense is assigned provisionally to the bestSense field of the corresponding DisambiguationWindowOccupant if, and only if, it has as yet had no bestSense assigned to it. If it already has a bestSense assigned to it, irrespective of whether it has already been in the target position, then a Paradox is recorded, that DisambiguationWindowOccupant is marked as paradoxical, and the existing bestSense is retained. If the target already has a bestSense assigned, then that bestSense is overwritten, but a Paradox is still recorded and the target is marked as paradoxical.

6.3.6.1.2 Marking the Disambiguation Output

After the target has been disambiguated, a DisambiguationOutputWord is created whose fields are the word field and the WordSense occupying the bestSense field from the DisambiguationWindowOccupant leaving the window, and Boolean fields, indicating whether the DisambiguationWindowOccupant was marked as paradoxical and whether its disambiguation as target defaulted to disambiguation by frequency (Fig. 10). The DisambiguationOutputWord is added to the output list.

[11] The selected senses are held temporarily in a List<WordSense> equal in size to the window, in which the target position is occupied by the selected sense of the target. That position in the list which corresponds to the other window occupant used in obtaining the highest score is occupied by the other selected sense. The remaining positions are occupied by nulls. This implementation facilitates compatibility with the B&P (§6.3.6.2) and Nearest Neighbours (§6.3.6.3) algorithms.

Once the whole text has been disambiguated, the output list is marked. Each `DisambiguationOutputWord` is passed to the `GoldStandardReader` for marking. If the `WordSense` stored in the `DisambiguationOutputWord` is null, or its POS does not match that of the corresponding `DisambiguationGoldStandardWord`, in which the `GoldStandardReader` holds the full information for the word represented by the `DisambiguationOutputWord`, it is marked as incorrect. A double check is made, that the sense number of the `WordSense` being marked is listed by the `DisambiguationGoldStandardWord` as a possible sense number and that the *lex_sense* component of the sense key encapsulated in the `WordSense` is also listed by the `DisambiguationGoldStandardWord`. If the results of these two checks conflict, the result from the sense number check overrides that of the sense key check[12], unless the lemma held in the `DisambiguationGoldStandardWord` differs from the word form of the `WordSense`, in which case it is marked as wrong.

In addition to marking each `DisambiguationOutputWord` right or wrong, the marking procedure also records the numbers of disambiguable words *W*, failures (no disambiguation result) *f*, defaults (where disambiguation reverted to disambiguation by frequency, but excluding failures) *d*, paradoxes (§6.3.6.1.1) *p*, correct non-defaults C_{-d} and correct defaults C_{+d}.

6.3.6.2 Differences between the One by One Generic Disambiguation Algorithm and Banerjee and Pedersen's Extended Gloss Overlaps

The generic algorithm described above differs in some important respects from Banerjee and Pedersen's (2002; 2003, §6.1.1.4) Extended Gloss Overlaps Algorithm. One obvious difference lies in the use of a range of morphological awareness levels (Tables 53-54). These must obviously be retained as the main objective is to compare disambiguation performance between them. However even when the *semantic* option is applied, which uses only WordNet relations, there are still important differences.

[12] Instances where this occurred were all found to be either lemma mismatches or errors in the encoding of sense keys in the gold standard dataset.

Fast Alternatives

Banerjee and Pedersen do not use 4 consecutive attempts at disambiguation with different measures, but only the method used in the fourth attempt (Gloss overlaps between all members of 2 collections of synsets). In order to perform experiments more comparable with theirs, only the fourth method is executed unless a *fast alternatives* option is adopted.

Asymmetrical Window at Each End

In order to have a constant number of words in the window for every target disambiguation, Banerjee and Pedersen (2002) use an asymmetrical window at the start and end of the text. The window is fully populated before disambiguation commences. The window is then frozen until all the words up to and including the one at the centre of the window have been disambiguated as targets, with reference to the same set of window occupants. Thereafter the window is advanced in the way described in §6.3.6.1 until the supply of text is exhausted, at which point the window is again frozen while the remaining words are disambiguated. This behaviour is reproduced in these WSD experiments by the *B&P Algorithm*, using a state machine.

Sense Combinations

Within the window, the generic algorithm described in §6.3.6.1 evaluates each pairing of the target with another word in the window, retaining only the best pairing of a target sense with another sense and the score from that best pairing. It then selects the best target sense from that pairing which produced the highest score.

Banerjee and Pedersen (2002), however, evaluate every possible combination of senses of the target word with senses of all the other words in the window, by adding the comparison scores of each pair within each combination, giving a total score for each combination. They then select the sense of the target word which occurs in that combination which has the highest score. This approach requires the retention of the target sense and score for every combination. The number of such combinations is given by

$$\prod_{i=1}^{w} S_i$$

where S_i is the number of senses of the word at position i and w is the window size. An order of magnitude approximation is given by

$$\left(\frac{\sum_{i=1}^{w} S_i}{w} \right)^w$$

This quickly leads to extreme demands on memory for window sizes > 3, but one might expect such a comprehensive set of comparisons to yield better results (but see §6.4.3).

In order to reproduce Banerjee and Pedersen's experiments as closely as possible, while keeping track of paradoxes, the B&P Algorithm has been implemented by associating each sense combination with a score each time the window is advanced. The score for each sense combination is calculated by adding together the scores for each combination of the target and another window occupant. The combination with the highest score is selected, from whose `WordSense` array the `bestSense` of the target is extracted and any paradoxes are recorded as in the One by One Algorithm (§6.3.6.1.1).

In order to speed up the disambiguation by avoiding repetitions of the same sense comparison, the pair of senses compared is stored with its score in a sense comparison map[13], so that if a comparison has already been made, its result can be retrieved instead of being recalculated. This optimisation is applicable to every disambiguation algorithm except Baseline[14].

[13] Class `SensePair` holds a score as well as a `WordSense` pair. Class `SenseComparisonMap`, houses a `Set<SensePair>` and a `Map<WordSense, Set<SensePair>>`, which enables navigation from any `WordSense` to any `SensePair` in which it participates. If `fastAlternatives` is true, one `SenseComparisonMap` is instantiated for use by each of the 4 consecutive disambiguation attempts. Each time the window is advanced, every `SensePair` mapped to be a sense of the `DisambiguationWindowOccupant` leaving the window is removed from the `SenseComparisonMap`.

[14] The One by One algorithm never uses sense combinations and requires a separate `SenseComparisonMap` for each combination of a `relatednessMeasure` and a light or heavy `measure` method, so that non-comparable scores do not get compared (§6.3.1).

6.3.6.3 Nearest Neighbours Algorithm

Because of the very high memory overhead of the B&P Algorithm (§6.3.6.2), it proved impossible to use it in experiments with any window size > 5. To address this, a compromise was sought between the One by One and B&P Algorithms. With window size 3, this compromise is identical to the B&P Algorithm, but with a larger window, the target and its immediate neighbours are treated as a sub-window for which a list of sense combinations is created to which the B&P Algorithm is applied. Another list of sense combinations is then created, from all those combinations of senses which include the *best* sense of the target as discovered by the application of the B&P Algorithm to the sub-window, but with *all* the senses of the target and *all* the senses of those occupants which were excluded from the sub-window, but are its immediate neighbours. The B&P Algorithm is then reapplied to the new list. This procedure is repeated until a best sense has been determined for every window occupant. The list returned by the last execution of the B&P Algorithm is then used as in §6.3.6.2. This method drastically reduces the maximum number of sense combinations that need to be stored at any one time. The storage requirement for the first application of the B&P Algorithm is given by

$$\prod_{i=1}^{3} S_i \quad \text{(§6.3.6.2)}$$

and the order of magnitude approximation is given by

$$\left(\frac{\sum_{i=1}^{3} S_i}{3} \right)^3$$

This requirement will not increase significantly with subsequent repetitions of the B&P Algorithm unless there are many more senses for the other words than for the members of the sub-window. This means that the Nearest Neighbours Disambiguation Algorithm can be successfully applied to larger windows, though it remains slow (§6.4.1).

6.3.6.4 Baseline Disambiguation by Frequency

The only other disambiguation algorithm used is Baseline Disambiguation by Frequency. This simply selects that `WordSense` from the possible senses of the target,

which has the highest Brown Corpus Frequency as recorded in WordNet. If more than one `WordSense` achieves the same highest frequency then a null `WordSense` is returned.

In addition to its application when gloss overlaps fail (§6.3.6.1.1), this simple measure has also been used as a control for all experiments, as in the SENSEVAL competitions (§6.2). Banerjee and Pedersen's (2002; 2003) failure to compare their results to this baseline, but only to a random selection baseline, is unfortunate.

6.4 Results

5 consecutive disambiguation runs were conducted, with the configurations listed in Table 54, using a variety of window sizes, but always including window sizes 3, 5 and 7, using each of the three algorithms, B&P, Nearest Neighbours and One by One (§6.3.6), on all three texts in the SENSEVAL-2 all words dataset. Some experiments were also conducted on SENSEVAL-3, but these were abandoned on account of the long execution times (§6.4.1). All algorithms were tested with the same parameter settings except for parameter `asymmetricalAtEnds`, which was true for B&P but false for the other algorithms. Lemmas were allowed, because the lemmas are encoded in the dataset and these sometimes bear no relation to the words for which they are proposed as lemmas, particularly in the case of proper nouns. Parts of speech were allowed, for consistency, because they have been allowed by Banerjee & Pedersen (2002; 2003). All algorithms were executed without the fast alternatives option, but the One by One Algorithm was subsequently re-run with this option (§6.4.3.4), which dramatically reduced execution time. As a control, the baseline disambiguation by frequency (§6.3.6.4), for which the window size is irrelevant was also run over the dataset.

6.4.1 Execution Times

The overall execution times and calculated words per second for each algorithm with window sizes 3, 5 and 7 are shown in Table 56 and are generally very slow, apart from baseline disambiguation by frequency and One by One with Fast Alternatives.

The execution times for One by One with Fast Alternatives are not comparable as experiments on the SEVSEVAL-3 dataset were dropped because of slow execution. The words per second figures are all comparable however, and show that the fast alternatives do save a great deal of time.

Table 56: WSD execution times

Algorithm	Dataset	Window size	HHH:MM:SS	Consec. configs.	Total words	Words per second
Baseline	Senseval2+3	n/a	000:03:18	1	4370	22.0707
B&P	Senseval2+3	3	147:03:56	5	21850	0.0413
		5	300:43:30	5	21850	0.0202
		7	Out of memory	5	21850	n/a
Nearest Neighbours	Senseval2+3	3	146:09:25	5	21850	0.0415
		5	316:23:17	5	21850	0.0192
		7	495:22:36	5	21850	0.0123
1X1	Senseval2+3	3	140:13:19	5	21850	0.0433
		5	312:18:29	5	21850	0.0194
		7	493:53:07	5	21850	0.0123
1X1 with fast alternatives	Senseval2	3	004:12:48	5	12105	0.7981
		5	008:37:00	5	12105	0.3902
		7	013:40:00	5	12105	0.2460

With the use of a sense comparison map to eliminate repeat calculations (§6.3.6.2), the mean number of gloss overlap calculations per word required for each configuration is large; an order of magnitude approximation is given by

$$\frac{wS_i^2r^2}{2}$$

where w is the window size, S_i is the mean number of senses per word and r is the mean number of relations in a `relativesList`. This approximation applies to every algorithm except Baseline and One by One with Fast Alternatives. There is little difference in execution times between the three main variants. The long execution times can be attributed partly to the overhead of the Java Virtual Machine. The inefficiency of the implementation of relations (§1.3.2.2 and footnote) undoubtedly also plays its part,

6.4.2 Performance Metrics

The performance metrics correspond to those set for the original SENSEVAL-2 evaluation exercise (§6.2.2). Recall R is represented by

$$R = \frac{C_{-d}}{W}$$

precision P is represented by

$$P = \frac{C_{-d}}{W - f - d}$$

and coverage **Cv** is represented by

$$C_v = \frac{w - f - d}{W}$$

where C_{-d} is the number of correct non-defaults, W is the number of words to be disambiguated, f is the number of failures and d is the number of defaults, excluding failures (§6.3.6.1.2).

For baseline disambiguation different metrics are required because all the non-failures are defaults:

$$R = \frac{C_{+d}}{W}$$

$$P = \frac{C_{+d}}{W - f}$$

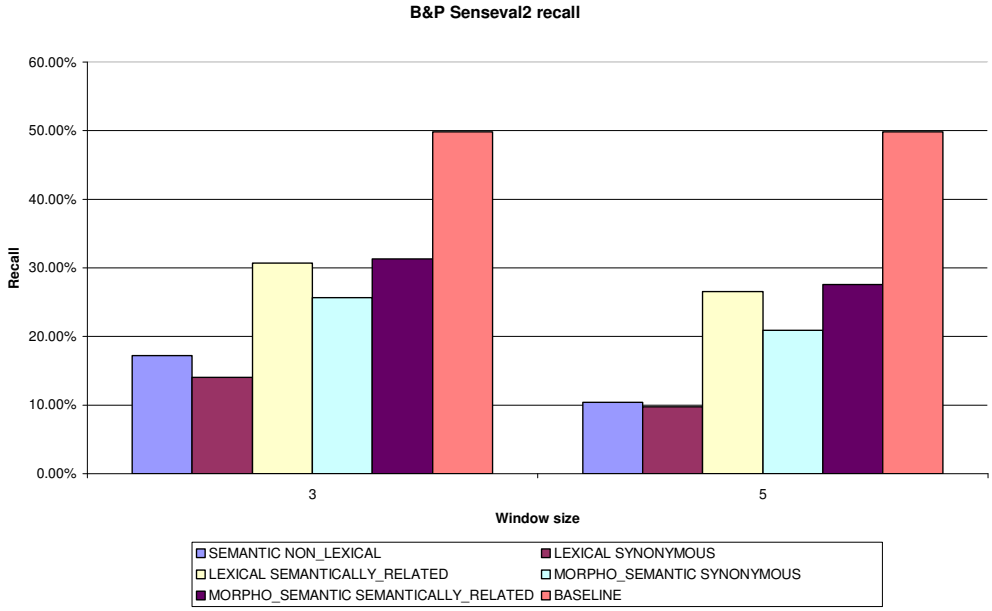
$$C_v = \frac{w - f}{W}$$

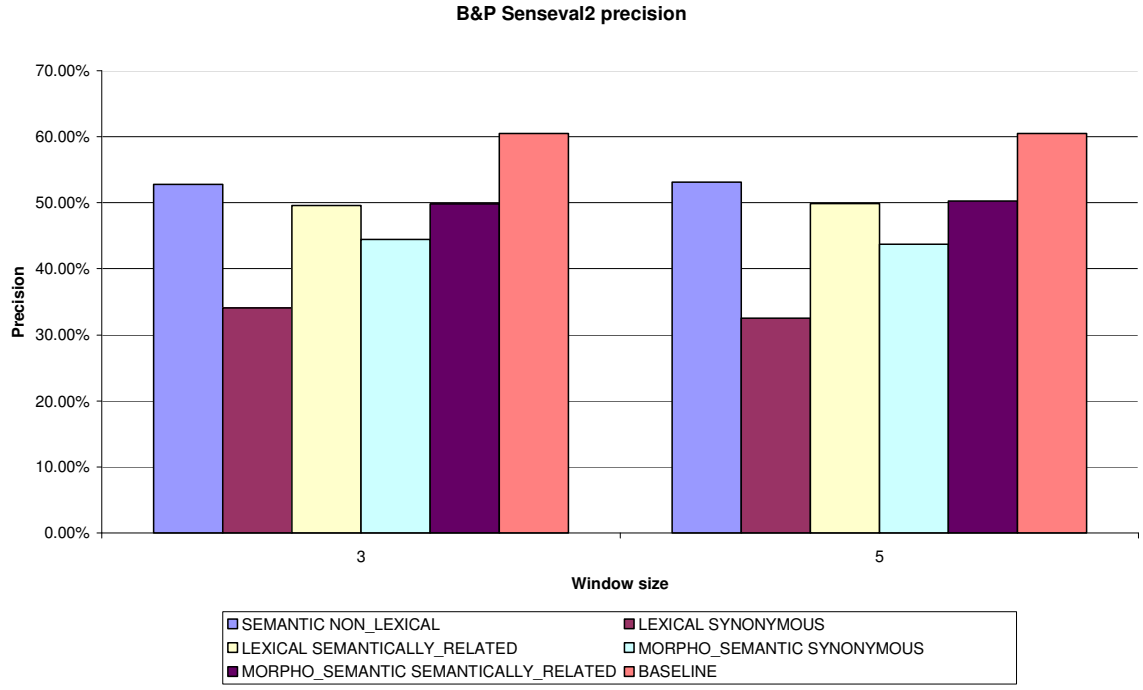
where C_{+d} is the number of correct defaults.

6.4.3 Performance

The results reported in this section are presented graphically; the underlying figures will be found in Appendix 63. The 5 different configurations used for testing each algorithm are referred to in the graphic legends in terms of their morphological awareness and lexical relativity (Table 54). These will be interpreted in the commentary in terms of the relations used.

Fig. 11: B&P WSD results





6.4.3.1 B&P Algorithm

The B&P Algorithm, which is implemented as closely as possible to the description by Banerjee & Pedersen (2002; 2003; §6.1.1.4), gave 17.22% recall and 52.78% precision (Fig. 11) with a window of size 3 and 10.37% recall and 53.18% precision

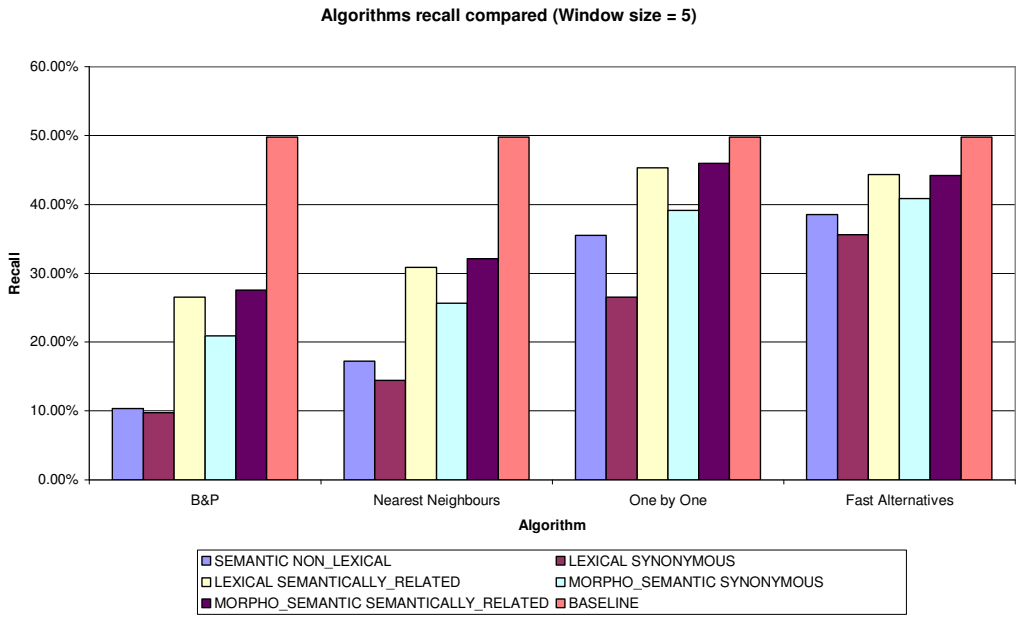
with a window of size 5, when applied using WordNet relations only. This compares with Banerjee & Pedersen's (2003) reported figures of 34.2% recall and 35.1% precision (§6.1.1.4). There are big disparities here. The principal known difference between the experimental setups is that Banerjee & Pedersen used the SENSEVAL-2 lexical samples task and the experiments described here used the all words task. It has been suggested that the all words task is more demanding than the lexical samples task (§6.2.2), which would account for the poor recall, but that doesn't explain why a much better precision has been achieved, nor why Banerjee & Pedersen's recall and precision figures are so close to each other while in the current experimental setup they are so far apart. The other main difference is in the modifications to WordNet discussed in §4, but it is not apparent why they should have these effects. One possible explanation for the disparities is a difference in behaviour when gloss overlaps do not identify a best sense for the target. The idea of defaulting to a frequency-based disambiguation was taken from Banerjee & Pedersen (2002), but seems to have been abandoned in Banerjee & Pedersen (2003). They may be allowing partial scores where the correct sense is among a set of identified best senses, whereas the methodology presented here defaults to a frequency-based disambiguation in those circumstances.

Banerjee & Pedersen neglect to compare their figures with the performance of a frequency-based algorithm. Their baseline is random sense selection, for which they report a recall and precision of 14.1%. The frequency-based baseline gives a recall of 49.81% and a precision of 60.48%, both of which exceed Banerjee & Pedersen's performance as well as the performance of the current version, not only when applied in a way as similar as possible to Banerjee & Pedersen's method, but also when using lexical relations, not only in this experiment but in all the others.

Surprisingly with the B&P Algorithm, recall is inferior with the larger window size, while precision barely changes at all. The recall of all configurations which use lexical relations (LEXICAL AND MORPHO−SEMANTIC), apart from the first (LEXICAL SYNONYMOUS) is significantly better than that achieved using WordNet relations alone (SEMANTIC NON−LEXICAL), while the precision achieved by using the lexical relations

of the WordNet relatives (SEMANTICALLY-RELATED) does not quite reach the precision achieved using WordNet relations alone.

Fig. 12: WSD algorithms compared (window size 5)



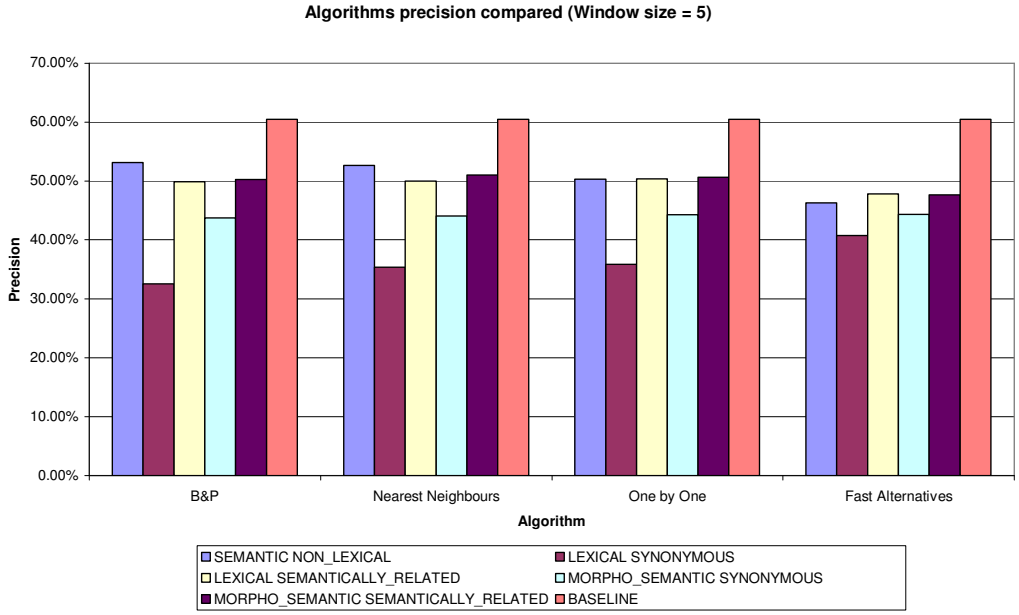
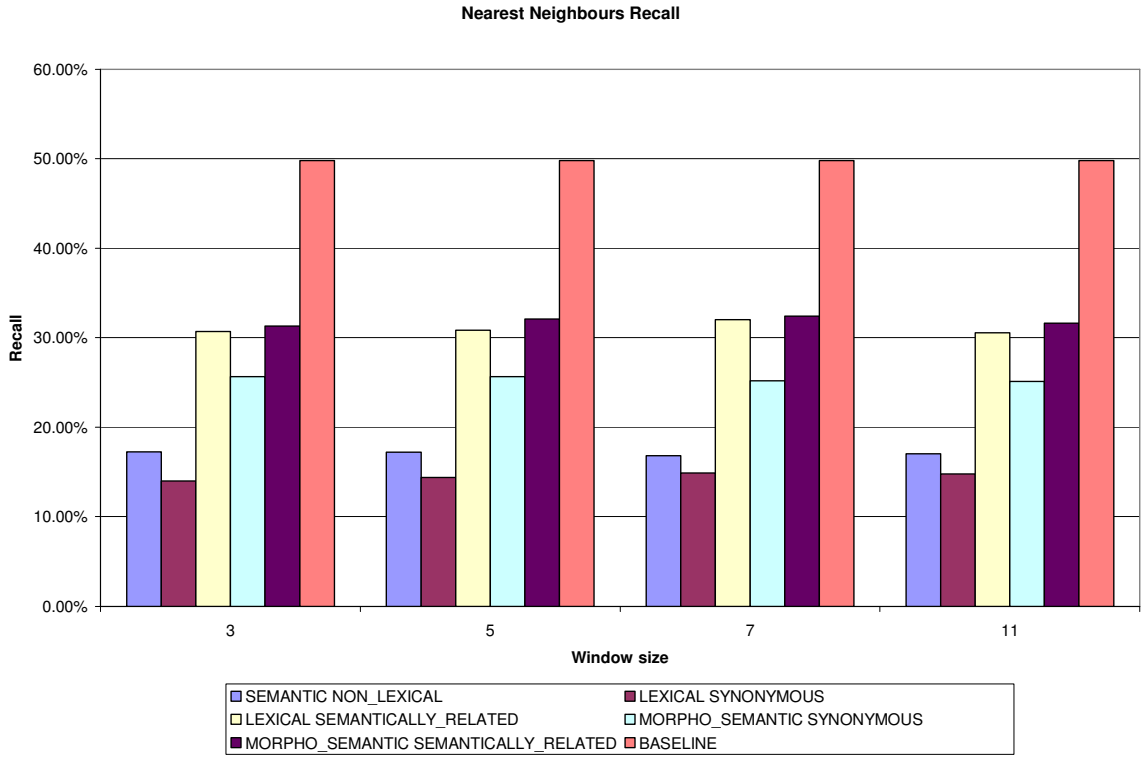
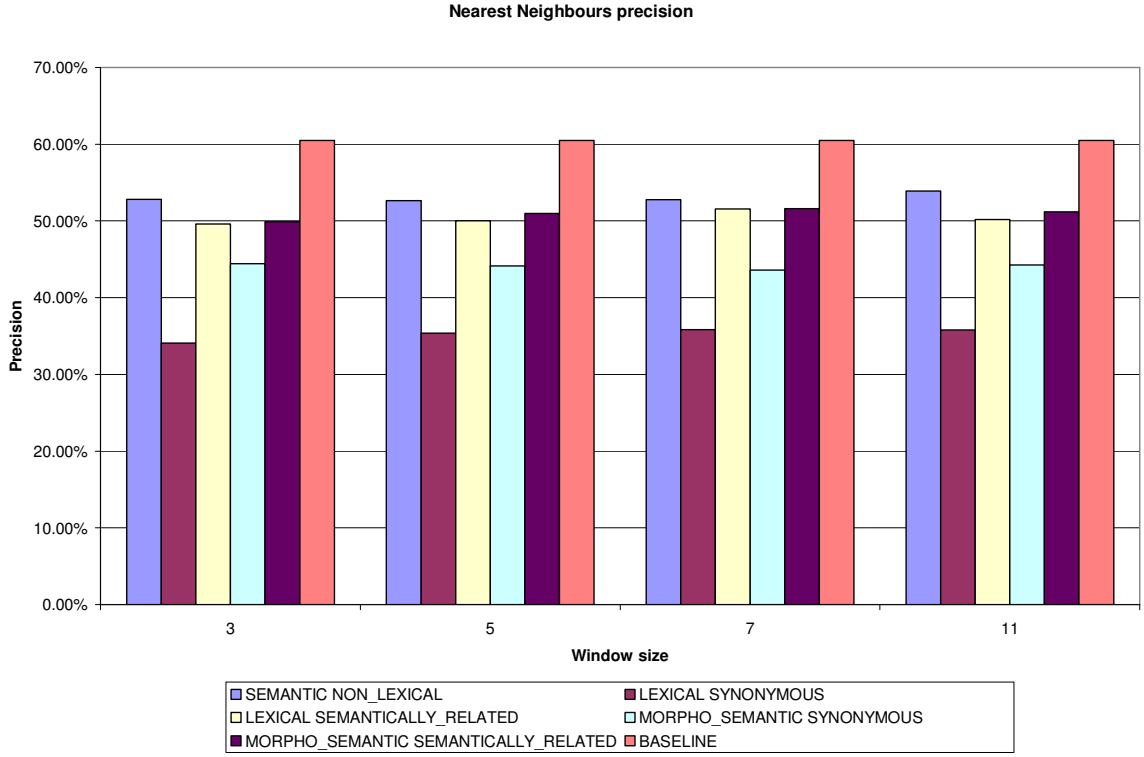


Fig. 13: Nearest Neighbours WSD results





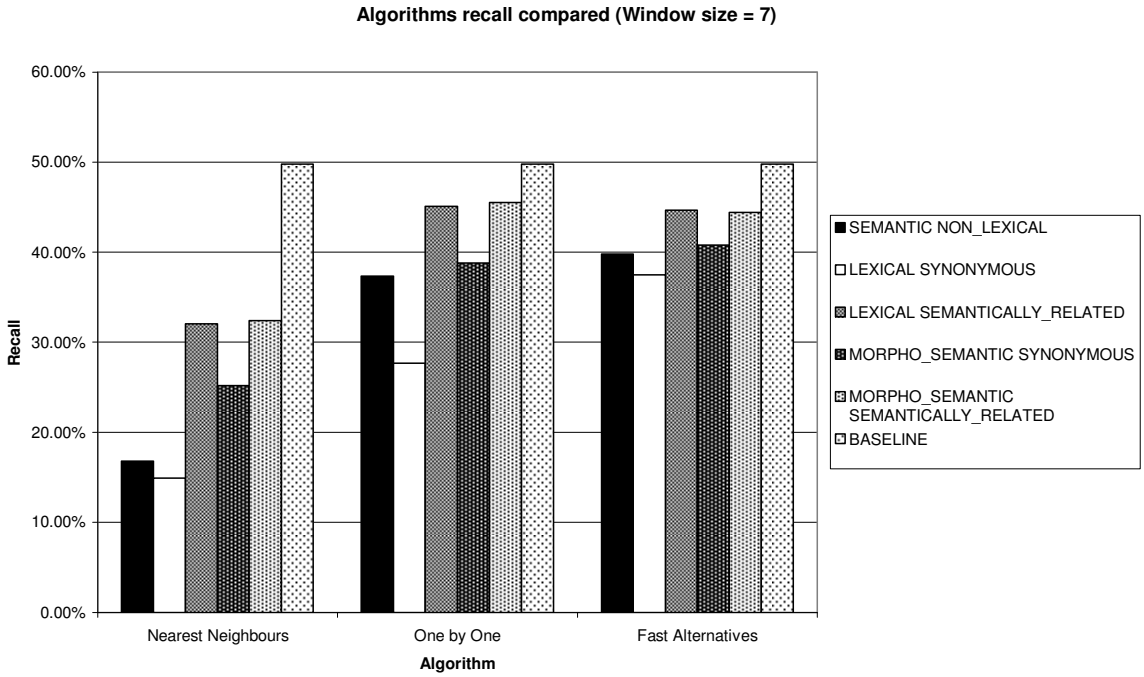
6.4.3.2 Nearest Neighbours Algorithm

The Nearest Neighbours Algorithm was devised because of the heavy memory requirements of the B&P Algorithm, such that it was impossible to complete experiments with a window size > 5. The Nearest Neighbours Algorithm behaves identically to the B&P Algorithm with window size 3. With window size 5 (Fig. 12), the Nearest Neighbours Algorithm gives significantly better recall all round; but the B&P Algorithm gives a slightly better precision using WordNet relations only (SEMANTIC NON-LEXICAL). Results from the Nearest Neighbours Algorithm are shown using window sizes 3, 5, 7 and 11. They show little variation with window size in either recall or precision (Fig. 13), though, when lexical relations are used (LEXICAL AND MORPHO-SEMANTIC), the best performance is achieved at window size 7. Recall is again much better using lexical relations, except for lexical relations of synonyms only (LEXICAL SYNONYMOUS).

6.4.3.3 One by One Algorithm

Unexpectedly, given that this is the least mathematically sophisticated algorithm, the One by One Algorithm gives significantly better recall than the Nearest Neighbours Algorithm irrespective of other variables (Figs. 12, 14, 15); but the Nearest Neighbours Algorithm gives a slightly better precision using WordNet relations only (SEMANTIC NON-LEXICAL), irrespective of window size, and with any configuration at window size 7. With this algorithm, using WordNet relations only loses its advantage over using lexical relations of WordNet relatives (SEMANTICALLY-RELATED), even when the WordNet relations themselves are excluded (LEXICAL SEMANTICALLY-RELATED), though using WordNet relations only (SEMANTIC NON-LEXICAL) gives slightly better precision with window size 3. The results from One by One show a significant improvement in recall with window size 5, when compared with window size 3, otherwise there is very little variance in performance with window size (Fig. 16). Recall is again much better using lexical relations (LEXICAL AND MORPHO-SEMANTIC), except for lexical relations of synonyms only (LEXICAL SYNONYMOUS).

Fig. 14: WSD algorithms compared (window size 7)



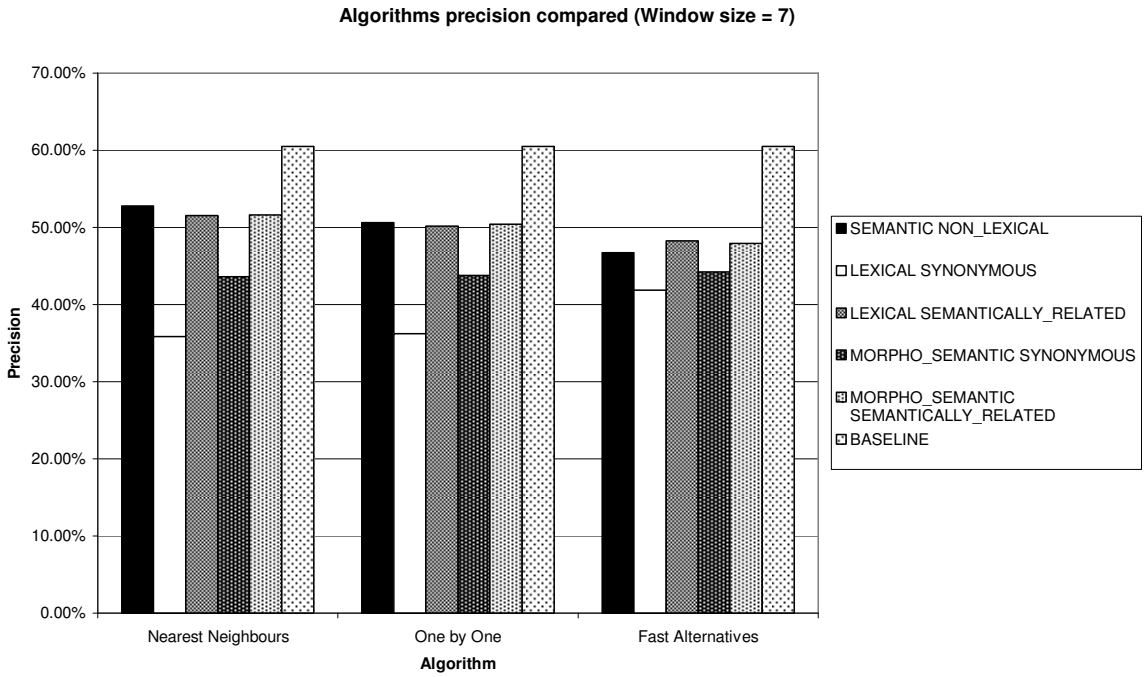
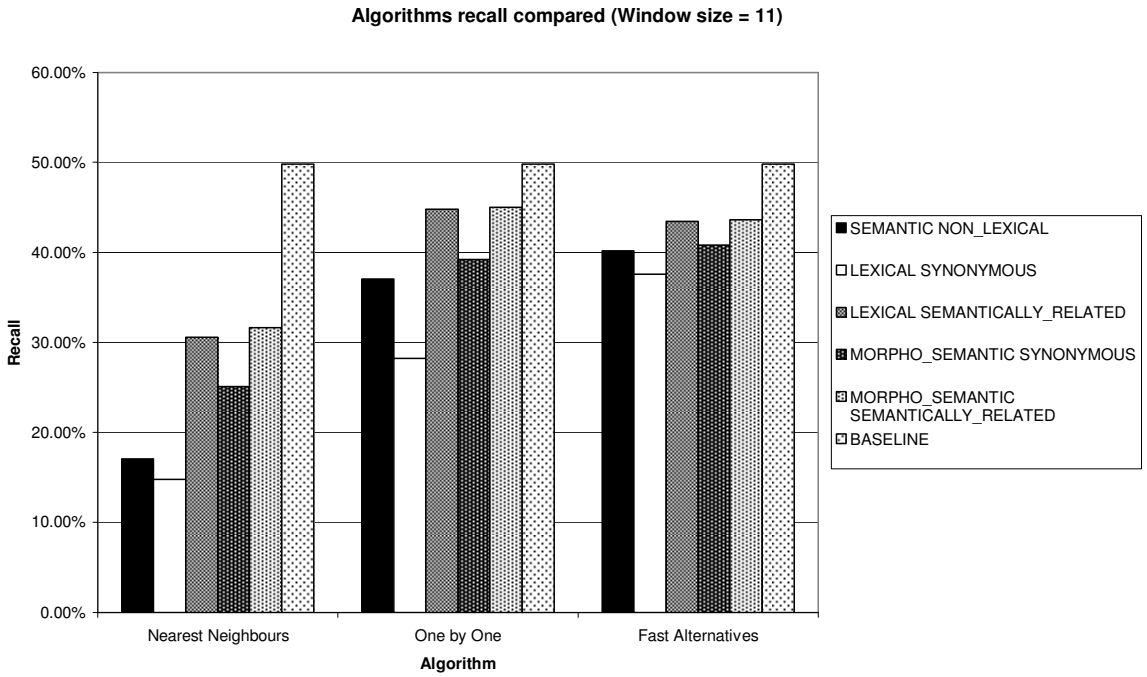


Fig. 15: WSD algorithms compared (window size 11)



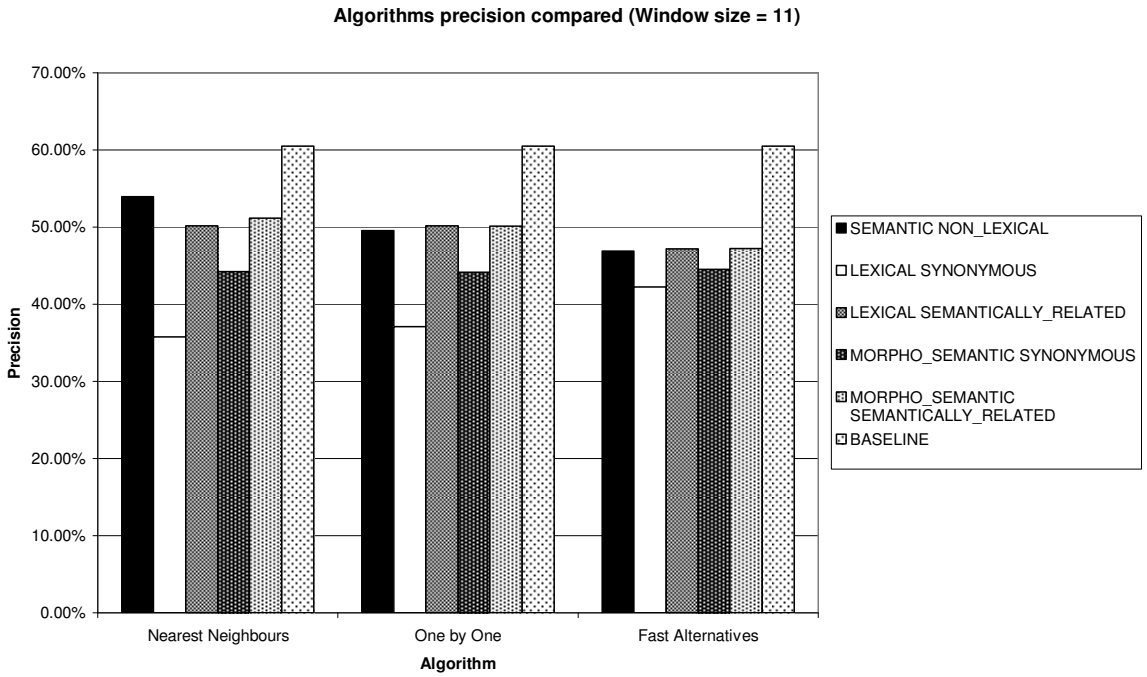
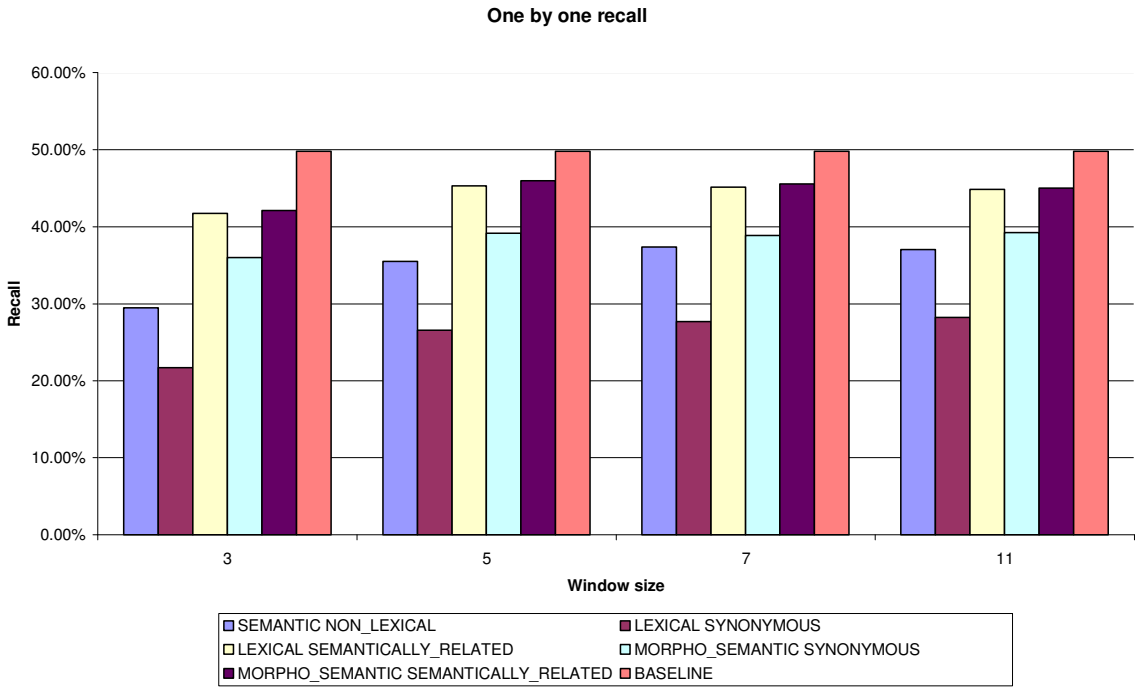


Fig. 16: One by One WSD results



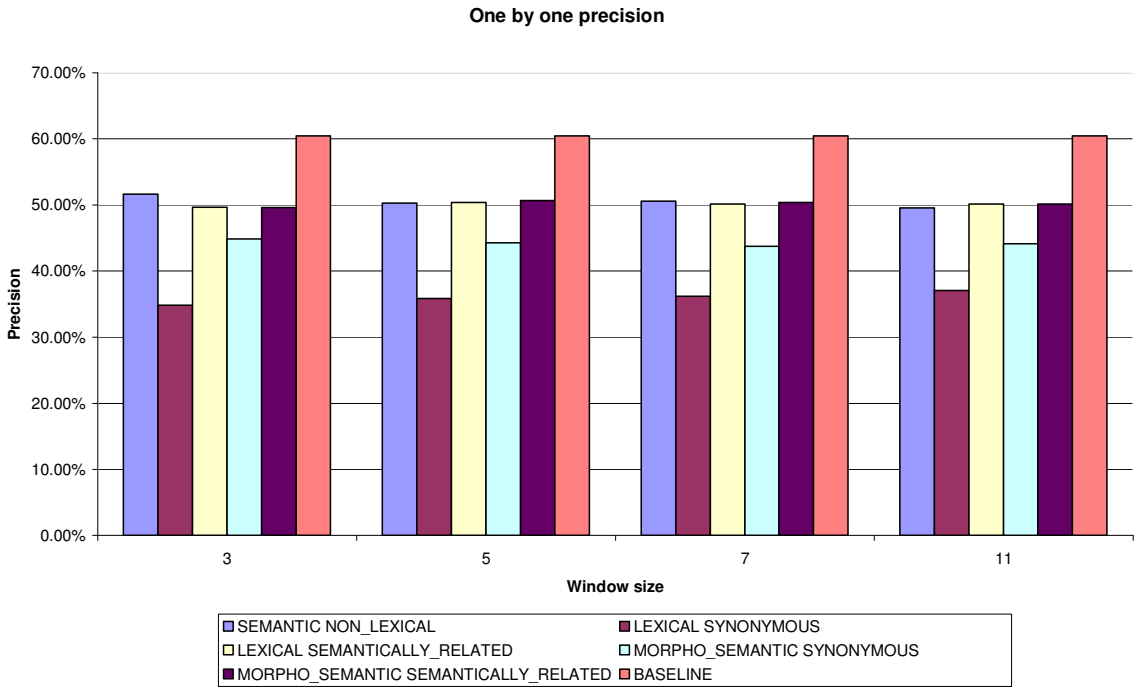
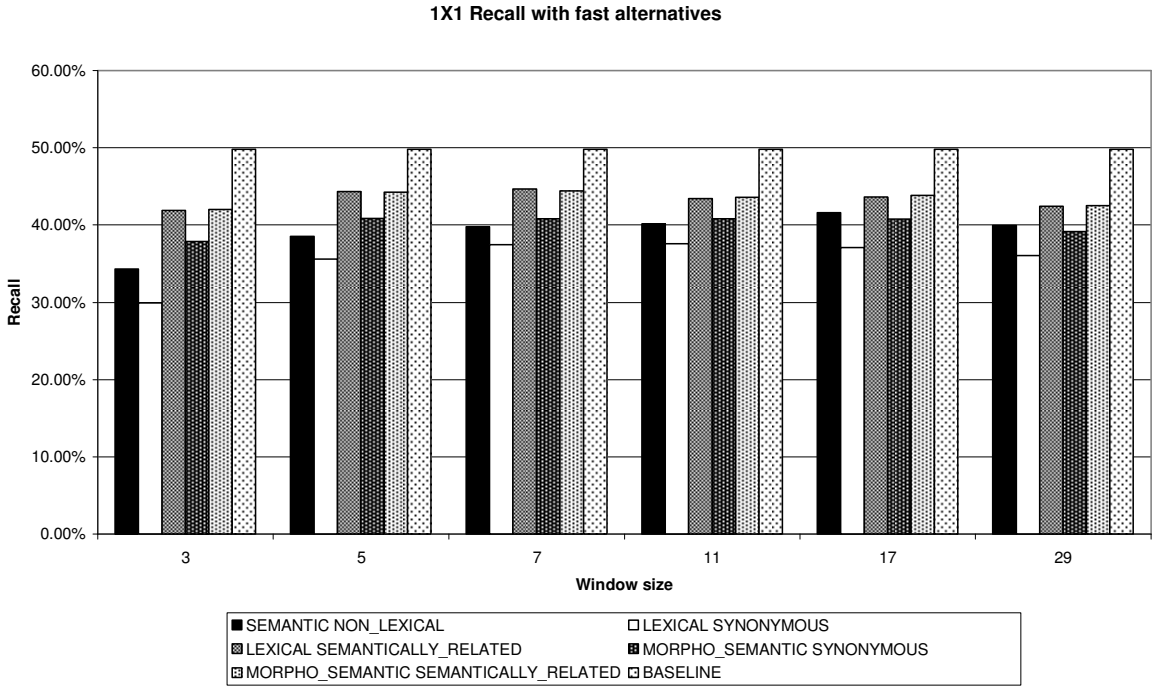
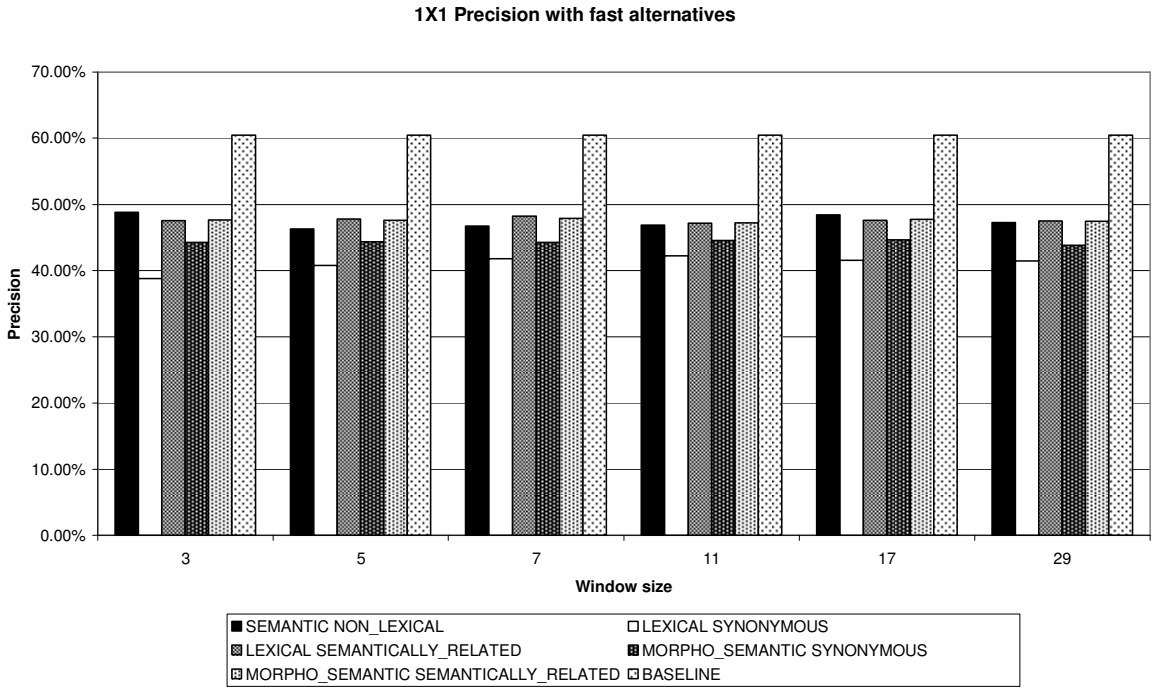


Fig. 17: One by One WSD results with fast alternatives





6.4.3.4 One by One Algorithm with Fast Alternatives

For a final test, the One by One Algorithm experiments were repeated with the fast alternatives option, which caused a dramatic improvement in execution speed (§6.4.1) at the price of a fall in precision (Figs. 12, 14, 15). The fall in precision did not however apply to configurations using the lexical relations of synonyms without WordNet relations (LEXICAL SYNONYMOUS), except at size 5. Recall improved for the otherwise worse recall configurations (SEMANTIC NON-LEXICAL, using WordNet relations only or LEXICAL SYNONYMOUS, using lexical relations of synonyms without WordNet relations).

Because of faster execution, results could be obtained using the One by One Algorithm with Fast Alternatives with larger window sizes (17 and 29 are shown; Fig. 17). Recall improves noticeably from size 3 to size 5 but then flattens out while precision also shows the greatest change between those window sizes, showing a noticeable fall between sizes 3 and 5 when using WordNet relations only (SEMANTIC NON-LEXICAL) and an improvement when using lexical relations of synonyms only (LEXICAL SYNONYMOUS), otherwise there is little variance with window size, though the optimum, when using lexical relations (LEXICAL AND MORPHO-SEMANTIC) seems to be around 11-17. The gap in recall between different configurations narrows as the window size increases with minimum variance around 11-17. Using WordNet relations only (SEMANTIC NON-LEXICAL) gives the best precision with window sizes 3 and 17; otherwise the best results are obtained from the lexical relations of the semantic relatives, with (MORPHO-SEMANTIC SEMANTICALLY-RELATED) or without (LEXICAL SEMANTICALLY-RELATED) the WordNet relations themselves.

6.4.4 Interpretation of Results

None of the results obtained from any of the evaluation experiments outperformed baseline disambiguation by frequency with respect to recall or precision. This does not reflect on the lexical relations as the failure applies whether they are used or not. It could be construed as reflecting on the gloss overlaps method. However the performance of the gloss overlaps method is dependent on the quality of the glosses,

which has been called into question (§2.3.1). The performance of Banerjee & Pedersen's extension to the gloss overlaps method (§6.1.1.4), incorporating WordNet relations clearly depends on the quality of the WordNet relations, which has also been seriously called into question (§2.2). While configurations which make more use of WordNet relations have generally performed better than others, this does not mean that a more consistent set of relations would not result in better performance. Doubts have also been raised about the SENSEVAL-2 dataset (§6.2.2) and indeed about the WordNet sense distinctions on which it is based (§2.1).

The best recall and a consistent level of precision are obtained using the lexical relations of the WordNet relatives, irrespective of which algorithm or which window size is used. The improvements to recall obtained by using lexical relations are not accompanied by a corresponding loss in precision. This fact alone endorses the usefulness of the lexical relations, which are all based on derivational morphology. It would be interesting to experiment with using more indirect lexical relations. With fast alternatives, variance in recall between the different configurations reduces as the window size is increased. Using WordNet relations only gives a slightly better precision with the B&P and Nearest Neighbours Algorithms, but only at window size 3 with One by One. Overall, configurations which use lexical relations outperform those which do not, though using only lexical relations of synonyms does not work as well as using only WordNet relations. These results demonstrate the utility of morphological enrichment, while reaffirming that of the WordNet relations.

There is surprisingly little variation with window size, the biggest variation being in recall between window sizes 3 and 5, where there is a noticeable improvement with the One by One Algorithm and a noticeable deterioration with B&P. Other variations with window size are too slight and inconsistent for any conclusions to be drawn from them.

Three different algorithms have been used for handling sense combinations, with the same underlying Extended Gloss Overlaps Disambiguation Algorithm. Of the three algorithms, One by One consistently gives the best recall and B&P gives the worst (Figs. 14 & 15). Even with fast alternatives, One by One still outperforms the others. Precision using WordNet relations only is best with B&P and worst with One by One,

but with any configuration using lexical relations these differences disappear. Since its advantage with respect to recall is much more than any disadvantage with respect to precision, one must conclude that One by One is the best algorithm, and that a more comprehensive comparison of sense combinations yields no advantage. The variant using fast alternatives offers a considerable advantage with regard to speed at the same time as an improvement in recall. It is arguable that these two factors outweigh any loss in precision.

All Lesk-based disambiguation algorithms are subject to paradoxes (§6.3.6.1.1), and the results show an abundance of these (Appendix 63). No analysis has yet been made of these, but their abundance does call the WordNet sense distinctions into question once again. Further research is needed to determine whether coarser sense distinctions, or mutual disambiguation (§6.3) can reduce the number of paradoxes and whether in so doing, it also improves the overall performance.

7 Conclusion and Further Research

This research project has demonstrated that it is possible, by a semi-supervised automatic process, to discover the morphological relations between words in a lexicon and their components and to enrich a lexicon with those relations. The semantic import of these relations can sometimes be defined as a relation type or lexical function (Vincze et al., 2008; §3.1.3), as typically between suffixations and their roots, but is often best represented by translation of morphemes such as prefixes and the stems to which affixes are applied. It also has been demonstrated that enrichment of a wordnet with morphological relations, to create a *morphosemantic wordnet,* can improve the performance of a disambiguation algorithm which measures semantic relatedness between word senses using the relations between them (§6). Thus it is clear that the enriched version of WordNet provides measurable benefits in linguistic analysis. Hence, the project aims (§1.2.3) have been achieved.

§7.1 summarises the utility and shortcomings of the WordNet model, the flaws identified in WordNet and recommendations for addressing them in future along with the reasons for the deployment of WordNet, despite the acknowledged flaws, explaining the immediate remedies adopted and emphasising the portability of the morphological analysis methodology to another lexical database. §7.2 reiterates some problems arising from previous research into morphological analysis and from the pilot study into a rule-based approach and how these problems were eventually addressed. §7.2 also recapitulates the main theoretical concepts arrived at and how they were implemented in the development of the morphological analyser. While some shortcomings are acknowledged, it is shown how a high level of precision was achieved through iterative development and evidence is provided to demonstrate the comprehensiveness of both the analysis and the enrichment. §7.3 outlines the requirements for using a morphologically enriched lexical database for WSD and draws conclusions from the disambiguation results, showing the utility of the morphosemantic wordnet created and how disappointing results reflect on Princeton WordNet. Attention is drawn to the advantages of the new variants of the Extended Gloss Overlaps Disambiguation Algorithm which have been developed. §7.4

summarises areas for further research including possible applications of derivational morphology, particularly in translation technology.

7.1 WordNet

Given the proposal for the morphological analysis and enrichment of WordNet and given an awareness of criticisms made of WordNet, it was considered necessary to investigate those criticisms to assess the suitability of WordNet for such analysis and enrichment.

The detailed investigation into WordNet (§2) would not have been possible without the creation of the open source object-oriented software model (§1.3). While the investigation into morphology (§3) could, for the most part, have been conducted without the model, clearly some lexicon was needed for the demonstration of the morphological analysis and enrichment methodology, and the lexicon used was provided by the model. While the methodology is portable to another lexicon, it would be impossible to test the usefulness of the morphological enrichment for WSD (§6) without a sense inventory. The WordNet word senses were used, despite their shortcomings (§2.1), because an entirely empirically based sense inventory was not available, though currently ongoing research (Hanks & Pustejowsky, 2005) may provide something approaching one. To deploy the WordNet word senses and the morphologically enriched lexicon for WSD clearly also depended on the use of the model.

Extensive use of the model has revealed some shortcomings of the software architecture. Its greatest weakness is the design of class `Relation`, where the target is not represented as a reference to the target object but as an integer representing a synset identifier, in the case of a `WordnetRelation`, with the addition of another integer representing a word number, in the case of a `WordSenseRelation`, or as a `String` representing a word or stem, in the case of a `LexicalRelation`. This architecture was employed to facilitate serialisation of the model but slows down the navigation of relations (§1.3.2.2 & note; §6.4.1). It would have been better to

represent targets as references and to devise a better serialisation algorithm. This will be addressed in any future version.

Turning now to the characteristics of WordNet itself, considerable doubt has been cast by contemporary corpus linguists and cognitive scientists upon the validity of the concept of a word sense (§2.1.1), which is the atomic concept in WordNet. The trend in modern lexicography is towards identifying senses in terms of usage. Lexicographic research in this area is ongoing (Hanks & Pustejowsky, 2005) and tends towards empirically founded distinctions with fine granularity.

Sense distinctions which are too fine (§2.1.2) create problems in NLP, increasing the need for disambiguation. The kinds of WSD needed for applications such as information retrieval and automatic translation are not necessarily the same: in the case of information retrieval, as with a search engine, a search term is often a single word with no collocates by which to disambiguate it; in the case of translation the kind of disambiguation required is into translation equivalents. The derivation of sense distinctions from translation equivalents found in parallel corpora (§2.1.1.3) is proposed as the way forward for the enumeration of word senses, and the resultant granularity is likely to be more tractable than one derived from monolingual collocation analysis, while the sense distinctions would be empirically based. There can never be any consensus as to the number of senses a word has as long as attempts to enumerate them approach the problem monolingually, because the boundaries between senses are necessarily fuzzy and new meaning extensions are constantly being devised, facts intimately related to linguistic creativity. This is an area where more research needs to be done. Meanwhile, within this project, the WordNet sense distinctions have necessarily been tolerated despite their inadequacy, an inadequacy reflected in the poor results from all the WSD tests, when compared to disambiguation by frequency (§6.4).

Consideration has been given to various proposals for clustering word senses or synsets (§2.1.2.3), but it became clear that the lexical clustering implicit in the lexicon provides the best foundation for encoding morphological relations (§3.5.3). Moreover, a methodology for the morphological enrichment of a lexicon has the advantage of being more portable to a better database, being clearly separable from WordNet. This

is not intended to imply that the implementation of a clustering algorithm to reduce wordnet granularity is not a worthwhile exercise

An essential feature of a wordnet is that, like a thesaurus, it provides a categorisation of meanings, frequently termed an ontology. A perfect ontology is impossible (§2.2.1) because it implies perfect world knowledge; all ontologies are bound to some set of philosophical assumptions. However there is no doubt that a formally constructed ontology is an improvement on an ad-hoc one such as WordNet's. Constructing a taxonomy by treating the main word in a gloss as the HYPERNYM of the word being defined is a valid approach but the results will only be as good as the glosses themselves, a prerequisite being that the glosses constitute formal definitions which comprise phrases which can be substituted for the words they define. This is often not the case, and with verbs it may not even be possible, as when a more particular verb requires a different preposition than a more general one. The online game approach of *jeux de mots* (§2.2.1.2) is the most empirical approach yet devised to the identification of the semantic relations which make up a lexical ontology. These and other approaches could all contribute to a better ontology. A comparison of the results from systematic application of these approaches would be a useful way forward.

There is some literature on the theoretical expectations of the verb taxonomy (§2.2.2.1) in a wordnet, but the investigation in §2.2.2 is the first time the WordNet taxonomy has been subjected to a systematic review in terms of those expectations, an exercise which could not have been performed without the prior construction of an object-oriented model. The investigation discovered an extremely wide divergence between theory and practice, and that standards being applied to the creation of other wordnets based on Princeton WordNet are much higher than those applied in the construction of Princeton WordNet itself.

To address the inconsistencies in the verb taxonomy, it is proposed that theoretical expectations of the inheritance of verb properties should be employed for a complete revision of the taxonomy. A prerequisite for such an endeavour is an adequate set of verb frames, to which the verbs are correctly matched. Investigation into the representation of verb syntax found that this was very far from being the case (§2.3). Not only is the set of verb frames inadequate, but the matching of verbs to frames is

erratic, both in terms of frames incorrectly assigned to verbs and correct frames not assigned. Syntactic uniformity across a synset has often been assumed where it does not apply, which in turn suggests that the allocation of verbs to synsets also needs re-examination. Some success has been achieved at redefining the verb frames by parsing usage examples (§2.3.2.3.4), but corpus validation of the results turned out to be too major a task to include within this research project and has been paused in order for the research presented in this thesis to be completed and presented, with the intention of completing it at the earliest opportunity.

Although the investigation into WordNet confirmed many criticisms and provoked more, in the absence of a freely available and equally comprehensive digital alternative, extensive use had to be made of it. The problem with WordNet lies not in its theoretical basis, but in the inconsistency between implementation and theory (§2.2-2.3). A suitable database could be constructed from a machine-readable dictionary, but that would be a research project in its own right and would be likely to inherit inconsistencies from the resource upon which it was based. These considerations confirmed the need for a lexicon-based methodology for the discovery and encoding of morphological relations which is portable to an empirically derived lexicon.

One problem which had to be faced was the presence in WordNet of only 4 out of 8 parts of speech. Prepositions (§4.2) are needed for the correct encoding of both verb syntax and derivational morphology, in particular the morphology of verbal phrases and the interpretation of prefixes. The addition of prepositions was made possible with the cooperation of the research team at The Preposition Project (§4.2.1.4). Adding pronouns would also be a big improvement to WordNet, though it was not relevant to the immediate research aims.

A preposition taxonomy was implemented, after learning from the problems with the verb taxonomy (§2.2.2). The Preposition Project's implicit taxonomy, based on digraph analysis and corroborated by semantic role analysis (§4.2.1), was used as a starting point, but it has been argued that a lexical taxonomy operates at a higher level. This has been implemented on top of the implicit taxonomy, using abstract synsets (§4.2.4).

Other improvements (§4.3) to the model were undertaken only insofar as they could be automated. The most important of these was the elimination of arbitrary encyclopaedic information in the encoding of proper nouns. This was done as much to make space for enrichment with lexical relations as in order to improve connectedness and reduce arbitrariness. This leaves a version of WordNet whose legacy imperfections are acknowledged but which can be used as a platform for morphological enrichment of the lexicon and for experiments to demonstrate the utility of that enrichment for improving WSD performance by a wordnet, irrespective of its inherited errors and inconsistencies. Because the morphological analyser applied to the lexicon is portable, it can be adapted to the analysis of any lexicon which satisfies the requirement that it differentiates between a minimum of eight parts of speech. Possession of corpus frequency data would be an advantage.

7.2 Morphological Analysis and Enrichment

A survey of recent publications calling for the morphological enrichment of WordNet (§3.1) showed a preference for rule-based approaches, without any serious attempt to implement such an approach, beyond the generalised spelling rules needed for stemming.

WordNet derivational pointers do not indicate the direction of derivation and only capture relatively few derivational phenomena (§§3.1.3, 3.2.2.4). A detailed investigation of the CatVar database (§3.1.2.1) found that it overgenerates and undergenerates, while its clusters of derivationally related words have no internal structure to show the direction of derivation, a problem addressed theoretically by the concept of a derivational tree (§3.1.4) and practically by enforcing a requirement in the software that every `LexicalRelation` specify the direction of derivation.

A systematic approach to the identification of morphological phenomena called for a theory-independent empirical approach to the algorithmic identification of morphological components. However the correct identification of the patterns of word formation in which these components participate called for the formulation of rules specifying relationships between morphemes and, as far as possible, the semantic

import of those relationships. This required some measure of human interpretation which needed to be based on linguistically informed observation.

The complete morphological analysis of the contents of the lexicon required the analysis of compound expressions and concatenations into their constituent words and the analysis of affixations into their constituent morphemes. The research undertaken has shown that a morphosemantic wordnet can be constructed by a hybrid approach (§3.5.4) combining the algorithmic identification of morphemes with rules governing their behaviour, to analyse, subject to minimal constraints, all truly non-atomic words in the lexicon iteratively into their components (§5). A morphological lexical database can be constructed from a lexicon without sense distinctions, while a morphosemantic wordnet requires sense distinctions and semantic relations.

A morphological rule represents a transformation between an input morpheme and an output morpheme either of which can be a null morpheme (where there is no affix). The significance of the transformation is expressed as a syntactic or semantic relation type (§3.2.2). As Fellbaum et al. (2007) reluctantly admit, there is no one-to-one mapping between *morphological* and *semantic* transformations. This problem has been addressed by the specification of more generic *syntactic* relation types (Appendix 22). Table 57 shows the distribution of relation types among type categories. The majority of root-derivative links[15] specify only the direction of derivation, typically because they have been determined algorithmically without reference to morphological rules, their semantic import generally being conveyed by a morpheme translation. Of the 18.25% of links where a semantic or syntactic relation type has been identified, all of which have been determined with reference to morphological rules, roughly two thirds are fully specified semantically. The remainder involve a syntactic transformation.

Morphological rules must be linguistically informed to minimise overgenerations of the kind found in CatVar (§3.1.2.1.2). This requires an understanding of the complex historical processes of word formation which have taken place in Latin and Anglo-Norman, best exemplified by the irregular behaviour of suffixes "-ion", "-ant" and

[15] *Derivational* type category.

Table 57: Distribution of relation types and lexical relations among relation type categories

Relation Type Category[16]	Types within this category		Links comprising ROOT-DERVATIVE pairs whose types belong to this category	
Semantic	51	60.00%	27055	12.37%
Syntactic	10	11.76%	11341	5.18%
Derivational	3	3.53%	178872	81.75%
Semantic/syntactic	10	11.76%	1534	0.70%
WordNet	11	12.94%	0	0.00%
TOTAL	85	100.00%	218802	100.00%

"-ent" (§3.2.2.1). English word formation processes are relatively simple by comparison. Given specialised knowledge about these processes, a provisional set of morphological rules could be formulated from a subset of the CatVar database (§3.2.2). Initial testing of the provisional ruleset (§3.2.2.2) showed overgeneration when applied to short words and where the application of multilingually formulated rules inadequately modelled Latin and Anglo-Norman word formation processes, but serious undergeneration arose where those word formation processes were not represented. Undergeneration also demonstrated that the process of morphological rule formulation would benefit from the input of empirical data from automatic suffix discovery (§3.4.2).

The problem of overgeneration when applying morphological rules to shorter words was addressed by specifying, for each rule, whether it is applicable to suffixation analysis when the output is monosyllabic (§5.1.1). The specification for each rule was kept under constant review in the light of overgenerations and undergenerations observed during iterative development. Undergeneration in the case of exceptions to th e specification of the applicability of rules to monosyllabic output was circumvented by allowing *reprieves* during secondary suffixation analysis (§5.3.14.2).

Some consideration was given to the possibility of using a Latin lexical resource to aid correct formulation of morphological rules to represent processes of Latin word formation, especially in relation to the "-ion" suffix which forms quasi-gerunds (§§3.5, 5.1.2). In the end, given a knowledge of Latin grammar, the alternative

[16] See Appendix 22.

approach of inference from co-occurrences of morphological patterns in the lexicon was preferred as quicker and easier to implement, but still required manual examination of a complete list of words ending in "-ion" which do not also end in "-ation" and similar lists for other suffixes. 213 new rules were added in this way to the original set of 147.

On the basis of observed undergeneration in the output, additional rules were formulated throughout the iterative development process, while in response to observed overgeneration, other rules were re-specified as multiple rules with longer suffixes. Altogether, a further 192 rules were added in the course of iterative development, bringing the total to 552.

A review of morphological analysis algorithms (§3.3) found that elementary spelling rules are ignored because of the common underlying *segmentation fallacy*, that morphological analysis can be performed reliably by word segmentation. In the hybrid model, the morphological rules apply character substitutions where necessary to avoid succumbing to this fallacy in the case of suffixations; when word-initial and word-terminal character sequences (*candidate affixes*) are collected into affix trees and counted by the Automatic Affix Discovery Algorithm (§3.4), it is not assumed that the residues from their removal (*stems*) are valid morphemes, and these stems do not feed directly into the morphological analysis.

There are two criteria for determining whether a candidate affix is a valid affix. The *duplication criterion* is easily assessed, but determination of whether a candidate affix satisfies the *semantic criterion* requires the deployment of heuristics. Several heuristics were applied successively to the output from automatic affix discovery to test their effectiveness at distinguishing meaningful from meaningless affixes. These heuristics presuppose the concepts of *affix frequency* (f_c) and *parent frequency* (f_p), where the parent of a prefix is the same prefix without the last character and the parent of a suffix is the suffix without the first character. Another relevant concept is the *stem validity quotient* (q_s) which represents that proportion of the stems, occurring with the same affix in different words, which is lexically valid. The heuristic

$$\frac{f_c^{\,2}}{f_p}$$ (§3.4.1.2),

has been referred to as the *default heuristic,* being the best performing heuristic which does not require q_s, adopted for the first experiments on automatic affix discovery. However, the heuristic

$$\frac{f_c^{\,2} q_s}{f_p}$$ (§3.4.4)

was subsequently found to perform better and so it was adopted for use in all phases of affixation analysis as the *optimal heuristic*, though the default heuristic has been retained as a control during iterative affixation analysis (§§5.3.14.3, 5.3.16).

The only advantage of the default heuristic over the optimal heuristic is its ability to distinguish between prefixations and concatenations. Automatic prefix discovery was originally applied experimentally to the entire lexicon, but in the context of the full morphological analysis of the lexicon, it has been applied to an atomic dictionary comprising only those words which have not already been analysed (§§5.3.3.1, 5.3.11.6). Before prefixation analysis begins, as many concatenations as possible have already been analysed and removed from the atomic dictionary. This removes any advantage the default heuristic might have. Similarly, the rhyming dictionary required by automatic suffix discovery was derived from the full lexicon for the initial experiments but is derived from the atomic dictionary for complete morphological analysis (§§5.3.3.2, 5.3.7.1).

The hybrid model includes the necessary *Root Identification Algorithm* (§5.2.2) to select which, if any, morphological rule to apply, given a suffix pre-identified by the output from automatic suffix discovery, and the *Word Analysis Algorithm* (§5.2.1), needed to analyse words manifesting a variety of morphological phenomena. The Word Analysis Algorithm was designed initially to perform concatenation analysis but developed into a generic algorithm, which is also used in prefixation analysis (§5.3.11), secondary suffixation analysis (§5.3.14) and stem analysis (§5.3.17.4). Its generic capability depends on the deployment of lists of candidate morphemes for the beginnings and ends of words, with a variable lexical validity requirement. The flexibility of this algorithm allowed extensive code re-use. Both algorithms were

developed iteratively in response to observed patterns of overgeneration and undergeneration.

Exceptions to lexical relationship patterns are a problem intrinsic to many languages, poorly handled by either a purely algorithmic approach (§3.3) or an over-rigid rule-based approach. The adoption of an iterative development process allowed the manual compilation of stoplists, to prevent the erroneous encoding of lexical relations where an exception applies. The stoplists function as feedback from the observation of erroneous results into the methods which produced those results. This feedback loop was applied to the initial results from many phases of morphological analysis, allowing 100% precision to be achieved. Homonym analysis with POS variation (§5.3.8) only achieves 92.6% precision for monosyllables because the monosyllabic output has not been subjected to this treatment. This extensive output would undoubtedly benefit from similar treatment. In the case of antonymous prefixations, the requirement for stoplists was reduced to a minimum by specifying morpheme exceptions and morpheme counter-exceptions (§5.3.5.2).

The concept of a prefix footprint (§3.2.2.3) assists in the identification of semantically identical irregular forms of common prefixes which have undergone *sandhi* modifications and need to be regularised. The concept of a linking vowel (§§3.2.2.3, 5.3.11.9) handles anomalies arising from collisions between prefixes which may or may not have a terminal vowel and stems which may or may not have an initial vowel. A distinction has been drawn (§5.3.11.1) between a known and finite set of irregular prefixes, which need to be identified from a footprint (§5.3.11.5), and an indeterminate set of regular prefixes, identified by automatic prefix discovery and subject to no spelling variations apart from linking vowel exceptions (§5.3.11.6). These concepts have allowed the segmentation fallacy to be avoided for a successful analysis of prefixations, which has not been attempted in either CatVar (§3.1.2) or WordNet (§3.1.3).

The successful implementation of prefixation analysis also depended on recognising fundamental differences between the properties of non-antonymous prefixations on the one hand, and common properties of suffixations and antonymous suffixations on the other. Unlike suffixes, prefixes, except where antonymous, do not lend themselves

to the formulation of morphological rules, because prefixations do not indicate the same kind of syntactic transformations as suffixations (§3.5). Words morphologically related through prefixation do not generally form multi-level morphological trees. Prefixations generally have dual inheritance from a prefix and a stem, whose semantic contributions can best be represented by translating them from their language of origin; a suffix by itself is, however, typically devoid of meaning until applied in a word, where its semantic contribution can be defined as a function, represented by the relation type of the morphological rule which holds between the suffix-bearing word and its parent in the derivational tree. In this respect also antonymous prefixations behave more like suffixations than other prefixations, except that the relation type represented is always ANTONYM. Consequently, morphological enrichment from non-antonymous prefixation analysis requires the encoding of two links, one between the prefixation and the meaning of the prefix and the other between the prefixation and the meaning of the stem (§5.3.11.7)[17], while morphological enrichment from suffixation or antonymous prefixation analysis requires only one link to be encoded, between the suffixation and its identified morphological root, specifying the relation type of the applicable morphological rule (§5.3.7.3), or between the antonymous prefixation and its root, specifying the ANTONYM relation type.

The recognition of the similarity between suffixations and antonymous prefixations and their differences from non-antonymous prefixations led to the productive intuition which gave rise to the *affix stripping precedence rule*, that antonymous prefix stripping takes precedence over suffix stripping which in turn takes precedence over non-antonymous prefix stripping (§3.5.1). This rule has been successfully adopted in morphological analysis. The few errors arising from exceptions to it were circumvented through the iterative development feedback loop. Precedence of concatenation analysis over affixation analysis was assumed (§§3.5.2, 5.3.4), but, because many affixes comprise character sequences identical to unrelated words (§5.3.4.2), this assumption caused massive overgeneration, to address which stoplists and startlists were deployed and three phases of concatenation analysis were interspersed with affixation analysis phases.

[17] In practice, the latter is implemented as an indirect relation via the stem itself, which is stored, unlike the prefix itself.

Morphological analysis and enrichment can proceed up to a certain point with a requirement that outputs be lexically valid (that they occur in the lexicon, as the specified POS, if any). The representation of the mechanics of suffix substitution by morphological rules allows this requirement to hold during primary suffixation analysis, and the requirement serves as a check on the validity of the analysis. Beyond this point, for prefixation analysis (§§5.3.11, 5.3.16) and secondary suffixation analysis (§5.3.14), because the analysis largely involves unravelling word formation processes which occurred in the context of other languages, the outputs (prefixes and stems) are often not lexically valid but are semantically valid. These word formation processes apply especially to scientific vocabulary. Scientists who are not also linguists could benefit from the translations of the prefixes and stems which have been used to convey their semantic content. *Prefixes* are not stored, because they are not subject to further analysis, and relations are encoded directly between prefixations and the corresponding prefix meanings. *Stems* are stored, for subsequent further analysis, in a stem dictionary. The decision not to store prefixes in a prefix dictionary, similar to the stem dictionary, was retrospectively unfortunate, in that it complicated the final stages of the analysis, in particular the recovery of original prefixations (§5.3.17.3.2).

In the absence of any control equivalent to a lexical validity requirement, the contents of the stem dictionary need to be treated with caution until it can be demonstrated that the semantic import of the stem is the same when it occurs in conjunction with any of its listed affixes. For this reason, stem interpretation (§5.3.17.3) requires significant manual intervention, and has been confined to stems which occur with at least 3 affixes.

Even when the analysis of words into their components has been completed, the morphological analysis is not complete as long as there are stems capable of being analysed further. To minimise the risk of errors, all phases of affixation analysis only allow the removal of one affix at a time, though primary suffixation analysis outputs words some of which are themselves suffixations analysed during the same phase. Consequently, secondary prefixes, and secondary suffixes associated with non-lexical stems, remain agglutinated to the stems. The purpose of stem analysis (§5.3.17.4) is to identify such affixations within the stem dictionary. Stem analysis is an innovative, fully automated procedure applied with a further modification of the Word Analysis

Algorithm. It discovers some lexically valid components (§5.3.17.4.4), to which the stem can be connected, as well as additional stems and prefix instances (§5.3.17.4.5). A more complete analysis of stems would require multilingual lexical resources. Stem analysis and reinterpretation bring the morphological analysis to its conclusion.

The *comprehensiveness of the morphological analysis* can be measured by examining the unanalysed words in the atomic dictionary. This includes some words (1.71% of the atomic dictionary samples; Table 46, §5.3.17) whose lexically valid roots have been omitted from WordNet and loan-words whose morphology belongs to exotic[18] languages (17.95%). Further analysis of the loan-words would also require multilingual resources, as they are mostly examples, unique in English, of foreign word formation patterns. There are also a few unusual affixations[19] (7.69%) which iterative affixation analysis (§§5.3.14.3, 5.3.16) has failed to capture. The secondary affix sets used during iterative affixation analysis contain character sequences, prioritised by heuristics because of their frequency, but which are semantically void, because the performance of the heuristics deteriorates as affixations are progressively removed from the atomic dictionary. These semantically void character sequences cannot be matched to morphological rules or prefix translations. The words in which they occur remain in the atomic dictionary and are recycled at each iteration. The size limitations placed on the secondary affix sets prevent unusual affixes from being represented because of this recycling. This could be addressed by increasing the size of the secondary affix sets or by preventing the recycling of invalid affixes. This would be likely to result in the successful analysis of up to 500 additional words, given that unusual affixations constitute roughly 7.7% of the atomic dictionary.

The *comprehensiveness of the morphological enrichment* can be measured by the number of lexical relations encoded in the lexicon. The results of the enrichment comprise 218802 links between words and their roots (other words and stems). Iterative development using stoplists ensured 100% precision from the main phases from which most of these links were created, namely primary concatenation analysis

[18] The term "exotic" here excludes the main ancestor languages of English (Anglo-Saxon, Anglo-Norman and Latin).
[19] e. g. "galactagogue", "logomach", "luminesce", "myxomycete", "neither", "pyelogram", "ritonavir", "vivisect".

(65% recall), primary suffixation analysis (98% recall) and primary prefixation analysis (96% recall).

7.3 Evaluation

While it would be possible to construct a lexical database entirely from morphological relations between words in a lexicon, this would not be a wordnet as generally understood and would not support WSD. As the morphological data encoded applies to words rather than word senses, it cannot contribute to WSD without reference to other data. WSD can only be performed when a set of senses of homonyms is provided. Moreover, while morphological relations have semantic import, there are many semantic relations which are not conveyed by morphology. For these reasons, the disambiguation experiments were conducted on the morphosemantic wordnet as a whole, rather than on its morphologically enriched lexicon component.

The utility of the morphosemantic wordnet was evaluated by comparing the disambiguation performance of a known algorithm which uses WordNet (*semantic*) relations with its performance when applied using morphological (*lexical*) relations and with its performance using both. The algorithm had to be one which uses only variables which are meaningful for both lexical and semantic relations (§6.1.1). The algorithm chosen was adapted from the Extended Gloss Overlaps Algorithm (§6.1.1.4) and performance was evaluated using the SENSEVAL-2 all words gold standard dataset (§6.2.2), using frequency-based disambiguation as a baseline.

Separate disambiguation experiments applied the lexical relations of the synonyms and the lexical relations of the semantic relatives (§6.3). Using the lexical relations of the semantic relatives in conjunction with the semantic relations themselves consistently improved recall when compared to using the semantic relations alone, demonstrating that morphological data contributes to WSD (§6.4). This clearly outweighed any corresponding loss of precision in a small number of experiments, demonstrating the utility of the morphological enrichment. The use of more indirect lexical relations might well lead to a further improvement.

The disambiguation experiments have also contributed better performing variants of Banerjee and Pedersen's (2002; 2003) Extended Gloss Overlaps Algorithm. Different high level algorithms were used for handling sense combinations, of which the simplest (One by One) consistently gave better recall than the memory-greedy B&P Algorithm, while the compromise Nearest Neighbours Algorithm consistently fell between the two. The B&P Algorithm gave better precision only when lexical relations were ignored. The original variant of the One by One Algorithm (One by One with Fast Alternatives), which only uses gloss overlaps where it cannot disambiguate using stronger sense match measures (§6.3.1), outperformed all the others and executes much more quickly. Little variation was found with window size, except that it became clear that a window size of 3 is too small.

The failure of any of the disambiguation experiments to outperform the baseline disambiguation by frequency (§6.4) clearly does not reflect on the utility of the morphological enrichment, since the enrichment improved performance. Rather it is a reflection on the quality of the WordNet sense distinctions, synonym identifications and semantic relations. These together determine the upper bound on the performance of any exercise which disambiguates into WordNet senses (§6.2) but, in combination with the glosses, they prevent any of the variants of the Extended Gloss Overlaps Algorithm from attaining even the lower bound (disambiguation by frequency), irrespective of whether morphological data is employed or not. This strongly suggests inconsistency between the glosses and the semantic relations.

7.4 Future Research Directions

Some possible improvements to the WordNet model have been identified which should be incorporated in any future version:

- revision of the software architecture of the WordNet model so as to facilitate faster navigation of relations (§1.3.2.2 & note);
- addition of pronouns to the WordNet model (§7.1).

A set of verb frames has been identified by parsing the usage examples of the WordNet verbal synsets, but attempts to validate this set against parsed sentences

from the BNC have not as yet been successful (§2.4). Completion of this work is a priority for the author and is a prerequisite for the revision of WordNet verb taxonomy and allocation of verbs to synsets in line with principles of verb frame inheritance (§2.3.2). The reorganisation of the rest of the taxonomy calls for a comparative evaluation of the results of systematic application of multiple approaches to ontology development (§7.1), possibly facilitated by the implementation of word sense / synset clustering according to a known clustering algorithm (§2.1.2.3). Ultimately, however, it might well be better to construct an entirely new wordnet from a machine-readable dictionary (§7.1) whose sense distinctions and glosses are consistent and demonstrably founded on empirical data. The author favours the definition of word senses from translation equivalents in parallel corpora over a monolingual approach which bases sense distinctions on usage patterns (§§2.1, 2.4) as being more likely to produce a finite set of discrete senses and more appropriate to applications in machine translation (§7.4.1).

Possible improvements to the morphological analyser have also been identified as follows:

- further investigation into the applicability of the semantic and syntactic types of identified morphological relations (§3.2);
- a review of the semantic correspondence between hyphenation components and the equivalent lexicon entries (§5.3.2.2 and note);
- modification of the homonym analysis phase with POS variation to employ a stoplist for monosyllables (§5.3.8);
- modification of the prefixation analysis phase to create a prefix dictionary, similar to the stem dictionary (§7.2);
- modification of the iterative affixation analysis phase to use larger secondary affix sets or to avoid recycling meaningless character combinations (§7.2);
- revision of the stoplist for tertiary concatenation analysis (§5.3.15);
- re-definition of class `POSTaggedStem` so that separate instances can be created of stems with the same orthography and POS (§5.3.17.3 and note);
- interpretation of stems occurring with fewer than 3 affixes (§5.3.17.3);
- translation of the information about morphological relations into a standard format (§5.3.18 and note).

It would be worthwhile to repeat the disambiguation experiments using more indirect lexical relations. It would also be interesting to see if better and less paradoxical disambiguation results could be obtained by applying mutual disambiguation techniques to a coarser-grained version of WordNet (§6.4.4) or by using the measures suggested by Hirst and St. Onge (1998; §6.1.1.2) and Sinha et al. (2006; §6.1.1.5).

The morphological analyser is intended to be portable. To demonstrate this portability, it needs to be applied to an alternative lexicon. A suitable lexicon has been derived from the BNC as a by-product of corpus parsing, but the prototype reveals the need for some improvements to the Lemmatiser component of the WordNet model (§1.3.2.5). Once the outstanding lemmatisation issues have been addressed, the alternative lexicon can be encoded in the same format as the main dictionary component of the WordNet-based lexicon, except without cross-referencing to the wordnet component. The morphological analyser can then be applied to it.

7.4.1 Applications of Derivational Morphology

The most obvious application of derivational morphology is in query processing, to find categorial variations (§3.1.2) on search terms, for instance to find a related verb or adjective when a query is expressed with a noun or for best-guessing what else a user might have meant by a lexically invalid search term. The methodology presented in this thesis can be used to produce more reliable categorial variation databases and extended to languages which do not possess any such database. Automatic affix discovery can be used to identify morphemes for which morphological rules need to be formulated for any language.

The morphological similarity between "geography" and "geology" is expressive of the common semantic domain to which these sciences apply. This illustrates how morphology could serve to inform the categorisation of words into semantic domains. This also has potential applications in query processing. The morphosemantic wordnet contains the necessary information.

Bilgin et al. (2004) suggest that morphological relations in one language can be used to discover semantic relations in another (§3.1.5). The relations discovered by the morphological analyser can be applied to lexical resources for other languages, and the adaptation of the analyser to such resources would allow further enrichment for English. If access to a wordnet for another language is not available, a translated wordnet could be created with the aid of a digital bilingual dictionary, along the lines suggested by de Melo & Weikum (2010). Such a wordnet would be inferior to a wordnet designed for the other language but might be sufficient for the discovery of morphological relations to translate as semantic relations.

WordNet has been used as a resource in Machine Translation (Langkilde & Knight, 1998). It is possible that the morphosemantic wordnet might perform better for this purpose. Habash (2002) describes an approach to machine translation, tailored to scenarios where there is a poverty of lexical resources for the source language but an abundance for the target language. The technique relies on overgeneration of possible translations followed by corpus-based statistical selection. The syntactic dependencies in the input are translated into thematic dependencies, from which alternative structural configurations are generated by reference to CatVar (§3.1.2). These are then realised syntactically before being passed to a statistical extractor which selects from the syntactic realisations by reference to corpus occurrences. This approach resolved 81% of a set of 48 translation divergences from Spanish to English. The results suggest that the combined analysis of syntax and morphology is useful for NLP tasks, but using a morphological database extracted from the morphosemantic wordnet would be an improvement on using CatVar.

The *quasi-gerunds*, ending in English with "-ion" and especially with "-tion" or "-ation" (§3.2.2.1) exist, often but not always with exactly the same meaning, in several European languages e. g.

- Latin Nominative -((a)t)io,
- Latin Genitive -((a)t)ionis,
- Italian -((a)z)ione,
- Spanish -((a)c)ión,
- Catalan -((a)c)ió,

- French -((a)t)ion,
- English -((a)t)ion.

The strong correlations between these quasi-gerunds in different languages has potential for economy in encoding interlingual lexical resources, inasmuch as exception lists to their correspondences in meaning, or "faux amis" (Rothwell, 1993), are likely to require much less storage than lexical entries associating them. The morphological rules which express the transformations involved between these quasi-gerunds in different languages are far more regular than the morphological rules which express the transformations between the quasi-gerunds and the corresponding verbs within each language. These considerations suggest that, even without any other semantic relations, a multilingual lexical database constructed entirely from morphological relations between words could be a useful resource, where the nodes hold word forms common to multiple languages and the arcs represent morphosemantic relations. Variations in meaning could be represented by language-specific morphosemantic relations or glosses. Alternatively, correlations between quasi-gerunds could serve as lynchpins, connecting ranges of related words between morphologically enriched lexical databases for individual languages.

Clearly a machine translation application did not fall within the scope of the research presented in this thesis. The author believes, however, that a morphologically enriched wordnet, whether based on improvements to WordNet as suggested, or entirely new and more empirically based (§7.4), could make a major contribution towards advances in this field. A monolingual morphosemantic wordnet could be deployed for the target language even where there is a poverty of resources for the source language, in the way outlined by Habash (2002), but the development of a multilingual morphosemantic wordnet, which could reduce redundancy and thereby economise on storage, could serve a more symmetric approach applicable to multiple languages. For related languages, this might eventually outperform existing approaches which ignore morphological data. While statistical machine translation has made great progress in recent times, syntactic and categorial variants still have a critical role to play in refining the output.

References

Amaro, R. 2006. WordNet as a base lexicon model for computation of verbal predicates. *3rd. Global WordNet Association Conference*, Jeju Island, Korea, 23-27 Jan. 2006.

Amaro, R., Chaves, R., Marrafa, P. & Mendes, S. 2006. Enriching Wordnets with new Relations and with Event and Argument Structures, *7th. International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, 19-25 Feb. 2006, 28-40.

Apresjan, J. 1973. Regular Polysemy, *Linguistics*, 142, 5-32.

Baldwin, T., Kordoni, R. & Villavicencio, A. 2009. Prepositions in Applications: A Survey and Introduction to the Special Issue, *Computational Linguistics*, 35, 2, 119-149.

Banerjee, S. & Pedersen, T. 2002. An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet, *3rd. International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, 17-23 Feb. 2002, 136-145.

Banerjee, S. & Pedersen, T. 2003. Extended Gloss Overlaps as a Measure of Semantic Relatedness, *International Joint Conference on Artificial Intelligence,* Acapulco, Mexico, 9-15 Aug. 2003.

Bilgin, O., Çetinoğlu, Ö. & Oflazer, K. 2004. Morphosemantic Relations In and Across Wordnets, *Proceedings of the 2nd. International WordNet Conference,* Brno, Czech Republic, 20-23 Jan. 2004, 60-66.

Blondin-Massé, A., Chicoisne, G., Gargouri, Y., Harnad, S., Picard, O. & Marcotte, O. 2008. How is Meaning Grounded in Dictionary Definitions, *Proceedings of 3rd. Textgraphs workshop on Graph-Based Algorithms in Natural Language Processing*, *22nd. International Conference on Computational Linguistics (COLING 2008),* Manchester, 18-22 Aug. 2008. 17-24.

Bosch, S., Fellbaum, C. & Pala, K. 2008. Enhancing WordNets with Morphological Relations: A Case Study from Czech, English and Zulu, *Proceedings of the 4th. Global WordNet Conference*, Szeged, Hungary, 22-25 Jan. 2008, 74-90.

Brewster, C., Iria, J., Ciravegna, F. & Wilks, Y. 2005. The Ontology: Chimaera or Pegasus, *Machine Learning for the Semantic Web Dagstuhl Seminar* 05071, Dagstuhl, Germany.

Budanitsky, A. & Hirst, G. 2004. Evaluating WordNet-based Measures in Lexical Semantic Relatedness, *Computational Linguistics*, 32, 13-47.

Burchfield, R. (Ed.). 1972. *A Supplement to the Oxford English Dictionary*, Oxford, Clarendon Press.

Chomsky, N. 1957. *Syntactic Structures*, The Hague, Mouton.

Crystal, D. 1980. *Dictionary of Linguistics and Phonetics*, Oxford, Blackwell.

Cunningham, H., Wilks, Y. & Gaizauskas, R. 1996. GATE - a general architecture for text engineering, *Proceedings of the 16th. Conference on Computational Linguistics*, Copenhagen, Denmark, 1057-1060.

de Melo, G. & Weikum, G. 2010. On the Utility of Automatically Generated WordNets, *5th. Global WordNet Association Conference*, Mumbai, India, 31 Jan.-4 Feb. 2010.

Dutoit, D. & Papadima, O. 2006. Alexandria as a Result of the Integration of WordNet and LDI, *Global WordNet Association Conference*, Jeju Island, Korea, 23-27 Jan. 2006.

Edmonds, P. & Cotton, S. 2001. SENSEVAL-2: Overview, Preiss & Yarowsky (eds.). *Proceedings of the Senseval2 Workshop*, Toulouse, France, Association for Computational Linguistics.

Edmonds, P. & Kilgarriff, A. 2002. Introduction to the special issue on evaluating word sense disambiguation systems, *Natural Language Engineering*, 8, 4, Cambridge University Press. 279-291.

EU. 2004. *Design and Development of a Multilingual Balkan Wordnet, BalkaNet, IST-2000-29388, WP8: Restructuring Improvement Word-Nets, Deliverable D8.1: Restructuring WordNets for the Balkan languages, Project Report*, European Commission.

Fellbaum, C. 1998. A Semantic Network of English Verbs, Fellbaum, C. (ed.). *WordNet: An Electronic Lexical Database*, Cambridge, MA., MIT Press, Chapter 3.

Fellbaum, C. & Miller, G. 2003. Morphosemantic links in WordNet, *Traitement automatique de langue*, 44, 2, 69-80.

Fellbaum, C., Osherson, A. & Clark, P. 2007. Putting Semantics into WordNet's "Morphosemantic" Links, *Proceedings of the 3rd. Language and Technology Conference*, Poznan, Poland, 5-7 Oct. 2007.

Fillmore, C. 1968. The Case for Case, Bach, E. & Harms R. (eds.). *Universals in Linguistic Theory*, New York, Holt, Rinehart & Wilson.

Goldsmith, J. 2001. Unsupervised Learning of the morphology of a Natural Language, *Computational Linguistics*, 27, 2, 153-198.

Graves, A. & Gutierrez, C. 2006. Data representations for WordNet: A Case for RDF, *3rd. Global WordNet Association Conference*, Jeju Island, Korea, 23-27 Jan. 2006.

Guarino, N. 1998. Some ontological principles for designing upper level lexical resources, *Proceedings of the 1st. International Conference on Language Resources*

and Evaluation (LREC 1998), Granada, Spain. Available from http://www.sciweavers.org/publications/some-ontological-principles-designing-upper-level-lexical-resources

Habash, N. 2002. Generation-Heavy Machine Translation, *Proceedings of the International Natural Language Generation Conference, Student Session*, New York, www.nizarhabash.com/publications/inlg-02.pdf.

Habash, N. & Dorr, B. 2003. A Categorial Variation Database for English, *Proceedings of the North American Association for Computational Linguistics*, Edmonton, Canada, 96-102. www.ldc.upenn.edu/acl/N/N03/N03-1013.pdf

Habash, N. & Dorr, B. No date. *System Demonstration - Catvar: A Database of Categorial Variations for English*, ftp://ftp.umiacs.umd.edu/pub/bonnie/catvar-demo.pdf.

Hafer, M. & Weiss, S. 1974. Word Segmentation by Letter Successor Varieties, *Information Storage and Retrieval*, 10, 371-385.

Hanks, P. 1997. Lexical Sets: Relevance and Probability, Lewandowska-Tomaszczyk, B. & Thelen M. (eds.). *Translation and Meaning, Part 4*, Maastricht, Netherlands, School of Translation and Interpreting.

Hanks, P. 2004. The Syntagmatics of Metaphor and Idiom, *International Journal of Lexicography*, 17, 3, 245-274.

Hanks, P. 2006. Conventions and Metaphors (Norms and Exploitations): *Why should Cognitive Linguists bother with Corpus Evidence? Cognitive-linguistic approaches: What can we gain by computational treatment of data?,* Theme session at DGKL-06/GCLA-06, Munich, Germany, 7 Oct. 2006.

Hanks, P. & Pustejovsky, J. 2005. A Pattern Dictionary for Natural Language Processing., *Revue Française de Linguistique Appliquée*, 10, 2, 63-82.

Harris, Z. 1955. From Phoneme to Morpheme, *Language*, 31, 190-222.

Hathout, N. 2008. Acquisition of the morphological similarity of the lexicon based on lexical similarity and formal analogy, *Proceedings of the 3rd. Textgraphs workshop on Graph-Based Algorithms in Natural Language Processing, at 22nd. International Conference on Computational Linguistics*, Manchester, 18-22 Aug. 2008, 1-8.

Hirst, G. & St-Onge, D. 1998. Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms, Fellbaum, C. (ed.). *WordNet: An Electronic Lexical Database*, Cambridge, MA., MIT Press, Chapter 13.

Huang, C-R., Tseng, I-J. & Dylan, B. 2002. Translating Lexical Semantic Relations: The 1st. Step Towards Multilingual Wordnets, *Proceedings of the 2002 SemaNet Workshop, 19th. International Conference on Computational Linguistics, COLING 2002*, 24 Aug.-1 Sep. 2002, Taipei, Taiwan.

Jackendoff, R. 1983. *Semantics and Cognition*, Cambridge, MA., MIT Press.

Jackendoff, R. 1990. *Semantic Structures*, Current Studies in Linguistics Series, Cambridge, MA., MIT Press.

Kahusk, N. 2010. Eurown: an EuroWordNet module for Python, *5th. Global WordNet Association Conference*, Mumbai, India, 31 Jan.-4 Feb. 2010.

Kilgarriff, A. 1997. I Don't Believe in WordSenses, *Computers and the Humanities*, 31, 91-113.

Kilgarriff, A. 1998a. Gold Standard Datasets for Evaluating Word Sense Disambiguation Programs, *Computer Speech and Language* 12, 4.

Kilgarriff, A. 1998b. SENSEVAL: An Exercise in Evaluating Word Sense Disambiguation Programs, *Proc. LREC*, Granada.

Kipper, K., Snyder, B. & Palmer, M. 2004. Using prepositions to extend a verb lexicon, *Proceedings of the HLT-NAACL 2004 Workshop on Computational Lexical Semantics*, Boston, MA, 23–29.

Knight, K. & Luk, S. 1994. Building a large-scale knowledge base for machine translation, *Proceedings of the 12th. national conference on Artificial intelligence*, Seattle, 773-778.

Koeva, S., Krstev, C. & Vitas, D. 2008. Morpho-semantic Relations in WordNet: A Case Study for two Slavic Languages, *Proceedings of the 4th. Global WordNet Conference*, Szeged, Hungary, 22-25 Jan. 2008, 239-253.

Kohl, K., Jones, D., Berwick, A. & Nomura, N. 1998. Representing Verb Alternations in WordNet, Fellbaum, C. (ed.). *WordNet: An Electronic Lexical Database*, Cambridge, MA., MIT Press, Chapter 6.

Kwon, H-S. 1997. *English Negative Prefixation: Past, Present and Future*, PhD. Thesis, School of English, Faculty of Arts, University of Birmingham.

Lakoff, G. 1987. *Women, Fire and Dangerous Things,* University of Chicago Press.

Langkilde, I. & Knight, K. 1998. Generation that Exploits Corpus-Based Statistical Knowledge, *Proceedings of the 36th. Annual Meeting of the Association for Computational Linguistics and the 17th. International Conference on Computational Linguistics (COLING-ACL '98)*. Montréal, Canada, 10-14 Aug. 1998, 704-710.

Leacock, C. & Chodorow, M. 1998. Combining Local context and WordNet Similarity for Word Sense Identification, Fellbaum, C. (ed.). *WordNet: An Electronic Lexical Database*, Cambridge, MA., MIT Press, Chapter 11.

Lee, E-R., Yoon, A-S. & Kwon H-C. 2006. Passive Verb Sense Distinction in Korean WordNet, *Global WordNet Association Conference*, Jeju Island, Korea, 23-27 Jan. 2006.

Lesk, M. 1986. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from a ice cream cone, *SIGDOC '86*, *Proceedings of the 5th. annual international conference on Systems documentation,* New York.

Levin, B. 1993. *English verb Classes and Alternations: A Preliminary Investigation*, University of Chicago Press.

Levine, J., Mason, T. & Brown, D. 1992. *Lex and Yacc*, Sebastopol, California, O'Reilly.

Liddell & Scott. 1889. *Intermediate Greek-English Lexicon founded upon Liddell & Scott's Greek-English Lexicon*, Oxford, Clarendon Press.

Litkowski, K. 2002. Digraph Analysis of Dictionary Preposition Definitions, *Proceedings of the SIGLEX / SENSEVAL Workshop on Word Sense Disambiguation, Recent Successes and Future Directions*, Philadelphia, 9-16 Jul. 2002, Association for Computational Linguistics.

Litkowski, K. & Hargraves, O. 2005. The Preposition Project, *Proceedings of the 2nd. ACL-SIGSEM Workshop on the Linguistic Dimensions of Prepositions and their Use in Computational Linguistics Formalisms and Applications*, 171-179, Colchester.

Litkowski, K. & Hargraves, O. 2006. Coverage and inheritance in the preposition project, *Proceedings of the 3rd. ACL-SIGSEM Workshop on Prepositions*, Trento, Italy, 37-44.

Litkowski, K. & Hargraves, O. 2007. SemEval-2007 task 06: Word-sense disambiguation of prepositions, *Proceedings of the 4th. International Workshop on Semantic Evaluations*, Prague, Czech Republic, 24-29.

Liu, Y., Jiangsheng, Y., Zhengshan, W. & Shiwen, Y. 2004. Two Kinds of Hypernymy Faults in WordNet: the Cases of Ring and Isolator, *Proceedings of the 2nd. Global WordNet Conference*, Brno, Czech Republic, 20-23 Jan. 2004. 347-351.

Marsh, L. & Goodman, G. 1925. *A Junior Course of English Grammar & Composition*, Blackie & Son Ltd.

Mbame, N. 2008. Towards a Morphodynamic WordNet of the Lexical Meaning, *Proceedings of the 4th. Global WordNet Conference*, Szeged, Hungary, 22-25 Jan. 2008, 304-310.

Mihalcea, R. & Moldovan, D. 2001. Automatic Generation of a Coarse Grained WordNet, *Proceedings of NAACL Workshop on WordNet and Other Lexical Resources*, Pittsburgh, PA, 2-7 Jun. 2001.

Miller, G. 1998. Nouns in WordNet, Fellbaum, C. (ed.). *WordNet, An Electronic Lexical Database*, Cambridge, MA., MIT Press, Chapter 1.

Miller & Johnson-Laird. 1976. *Language & Perception,* Cambridge University Press, Chapter 7.

Minnen, G., Carroll, J. & Pearce, D. 2001. Applied morphological processing of English, *Natural Language Engineering*, 7, 3, 207-223.

Moens, M. & Steedman, M. 1998. Temporal Ontology and Temporal Reference, *Computational Linguistics*, 14, 2, 15-28.

NODE. 1998. Pearsall, J. (ed.). *The New Oxford Dictionary of English*, Oxford, Clarendon Press.

ODE. 2003. Stevenson, A. & Soanes, C. (eds.). *The Oxford Dictionary of English*, Oxford, Clarendon Press.

OED1. 1971-80. *The Compact Edition of the Oxford English Dictionary, Complete Text Reproduced Micrographically*, Oxford University Press.

OED2. 2001-2010. Harper, D. *Online Etymology Dictionary*, http://www.etymonline.com/

Onions, C. (Ed.). 1966. *The Oxford Dictionary of English Etymology*, Oxford, Clarendon Press.

Pala, K. & Smrž, P. 2004. Top Ontology as a Tool for Semantic Role Tagging, *4th. International Conference on Language Resources and Evaluation (LREC 2004)*, Lisbon, 26-28 May 2004, 1897-1900.

Peters, W., Peters, I. & Vossen, P. 1998. Automatic Sense Clustering in EuroWordNet, *Proceedings of the 1st. Conference on Language Resources and Evaluation (LREC 1998)*, Granada, Spain, 28-30 May 1998.

Poesio, M., Ishikawa, T., im Walde, S. & Vieira, R. 2002. Acquiring Lexical Knowledge For Anaphora Resolution, *Proceedings of the 3rd. Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Spain.

Porter, M. 1980. An algorithm for suffix stripping, *Program*, 14, 3, 130-137.

Pustejovsky, J. 1991. The Generative Lexicon, *Computational Linguistics*, 17, 4. 409-441.

Pustejovsky, J. 1995. *The Generative Lexicon*, Cambridge, MA., MIT Press.

Quirk, R., Greenbaum, S., Leech, G. & Svartik, J. 1985. *A Comprehensive Grammar of the English language*, London, Longman.

Resnik, P. & Yarowsky, D. 1997. A perspective on word sense disambiguation methods and their evaluation, Light, M. (ed.) *Tagging Text with Lexical Semantics: Why, what and How?*, Washington, SIGLEX (Lexicon Special Interest Group), ACL, 79-86.

Richens, T. 2008. Anomalies in the WordNet Verb Hierarchy, *Proceedings of the 22nd. International Conference on Computational Linguistics (COLING 2008)*, Manchester, 18-22 Aug. 2008.

Richens, T. 2009a. Linguistically-Motivated Automatic Morphological Analysis for Wordnet Enrichment*, 6th. International Workshop on Natural Language Processing and Cognitive Science (NLPCS 2009) in conjunction with ICEIS 2009*, Milan, Italy, 6-7 May 2009, 36-45.

Richens, T. 2009b. Automatic Affix Discovery, *Human Language Technologies as a Challenge for Computer Science and Linguistics, Proceeedings of the 4th. Language & Technology Conference*, 6-8 Nov. 2009, Poznań, Poland, 462-6.

Rothwell, W. 1993. The Legacy of Anglo-French: *Faux amis* in French and English, *Zeitschrift für romanische Philologie,* 109, Max Niemeyer Verlag, 16-46.

Ruppenhofer, J., Miriam, M., Petruck, R., Johnson, C. & Scheffczyk. 2006. *FrameNet II: Extended theory and Practice*, Berkeley FrameNet.

Sagot, B. & Fišer, D. 2008. Building a free French wordnet from multilingual resources, *OntoLex 2008*, Marrakech, Morroco.

Simpson, D. 1966. *Cassell's New Latin Dictionary (4th. Edition)*, London, Cassell.

Sinha, M., Reddy, M. & Bhattacharaya, P. 2006. An approach towards Construction and Application of a Multilingual Indo-WordNet, *3rd. Global WordNet Association Conference*, Jeju Island, Korea, 23-27 Jan. 2006.

Smrž, P. 2004. Quality Control for Wordnet Development, *Proceedings of the 2nd. Global WordNet Conference,* Brno, Czech Republic, 20-23 Jan. 2004, 206-212.

Stetina, J. & Nagao, M. 1997. Corpus Based PP Attachment Ambiguity Resolution with a Semantic Dictionary, Zhou, J. & Church, K. (Eds.). *Proceedings of the 5th. Workshop on Very Large Corpora (WAC5*), 18-20 Aug.1997, Beijing and Honk Kong, 66-80.

Trautwein, M. & Grenon, P. 2004. *Proceedings of the 2nd. International WordNet Conference,* Brno, Czech Republic, 20-23 Jan. 2004, 341-346.

Veale, T. 2006. A Typology of Lexical Analogy in WordNet, *Global WordNet Association Conference*, Jeju Island, Korea, 23-27 Jan. 2006, 105-110.

Vendler, Z. 1967. *Linguistics in Philosophy*, Ithaca & London, Cornell University Press, Chapter 4.

Verspoor, C. 1997. *Contextually-Dependent Lexical Semantics*, PhD. Dissertation, Edinburgh.

Vincze, V., Almási, A. & Szauter, D. 2008. Comparing WordNet Relations to Lexical Functions, *Proceedings of the 4th. Global WordNet Conference*, Szeged, Hungary, 22-25 Jan. 2008, 462-473.

Vossen, P. (ed.). 2002. *EuroWordNet, General Document*, http://www.vossen.info/docs/2002/EWNGeneral.pdf.

Vossen, P. 2004. EuroWordNet: a multilingual database of autonomous and language-specific wordnets connected via an Inter-Lingual-Index, *International Journal of Lexicography*, 17, 2, 161-173.

Wong, S. 2004. Fighting Arbitrariness in WordNet-like Lexical Databases. A Natural Language Motivated Remedy, *Proceedings of the 2nd. International WordNet Conference,* Brno, Czech Republic, 20-23 Jan. 2004, 234-241.

URLs of Digital Resources

Aston Corpus Network http://acorn.aston.ac.uk/

British National Corpus: http://www.natcorp.ox.ac.uk/

Cambridge Advanced Learner's Dictionary online: http://dictionary.cambridge.org/

CatVar: http://clipdemos.umiacs.umd.edu/catvar/

FrameNet: http://framenet.icsi.berkeley.edu/

Jeux de Mots: http://www.lirmm.fr/jeuxdemots/

Online Etymology Dictionary: http://www.etymonline.com/

Perseus: http://www.perseus.tufts.edu/

Propbank: http://verbs.colorado.edu/~mpalmer/projects/ace.html

The Preposition Project: http://www.clres.com/prepositions.html

SEMCOR version of SENSEVAL-2: http://www.cse.unt.edu/~rada/downloads.html

Stanford Parser: http://nlp.stanford.edu/software/lex-parser.shtml

Trésor de la Langue Française: http://atilf.atilf.fr/

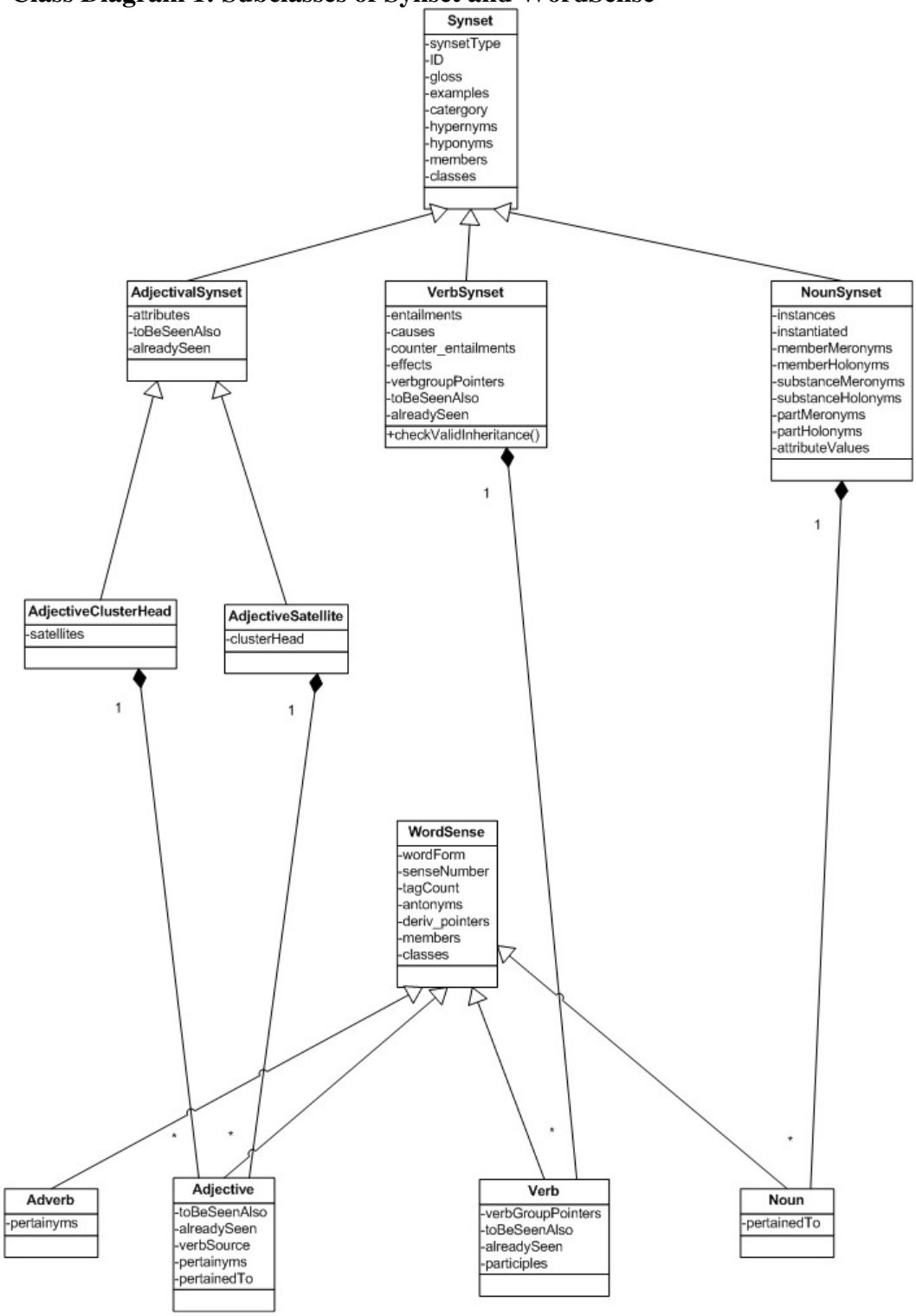
VerbNet: http://verbs.colorado.edu/~mpalmer/projects/verbnet.html

WordNet: http://wordnet.princeton.edu/

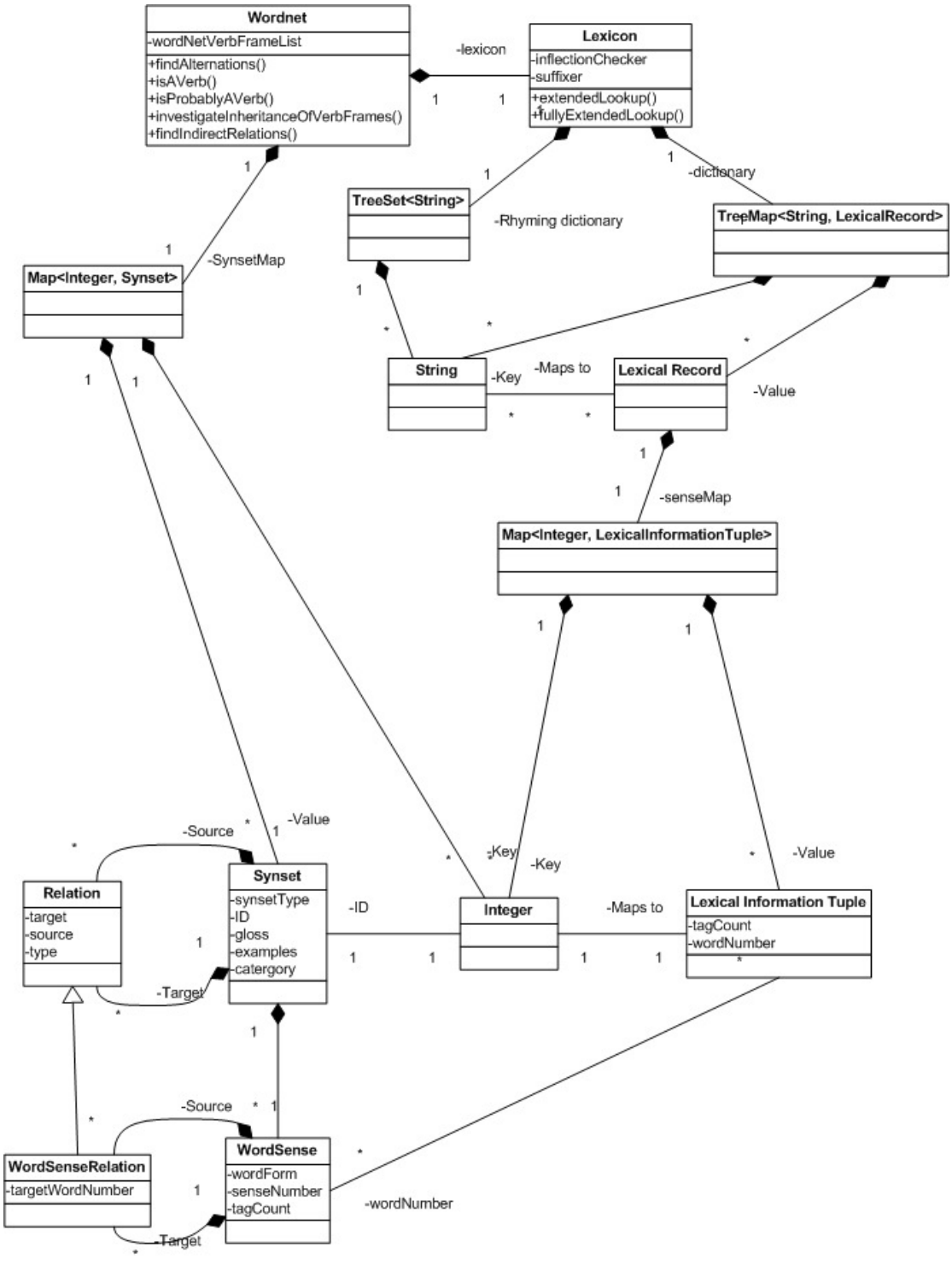
Class Diagrams

(*only selected fields and methods referred to are shown*)

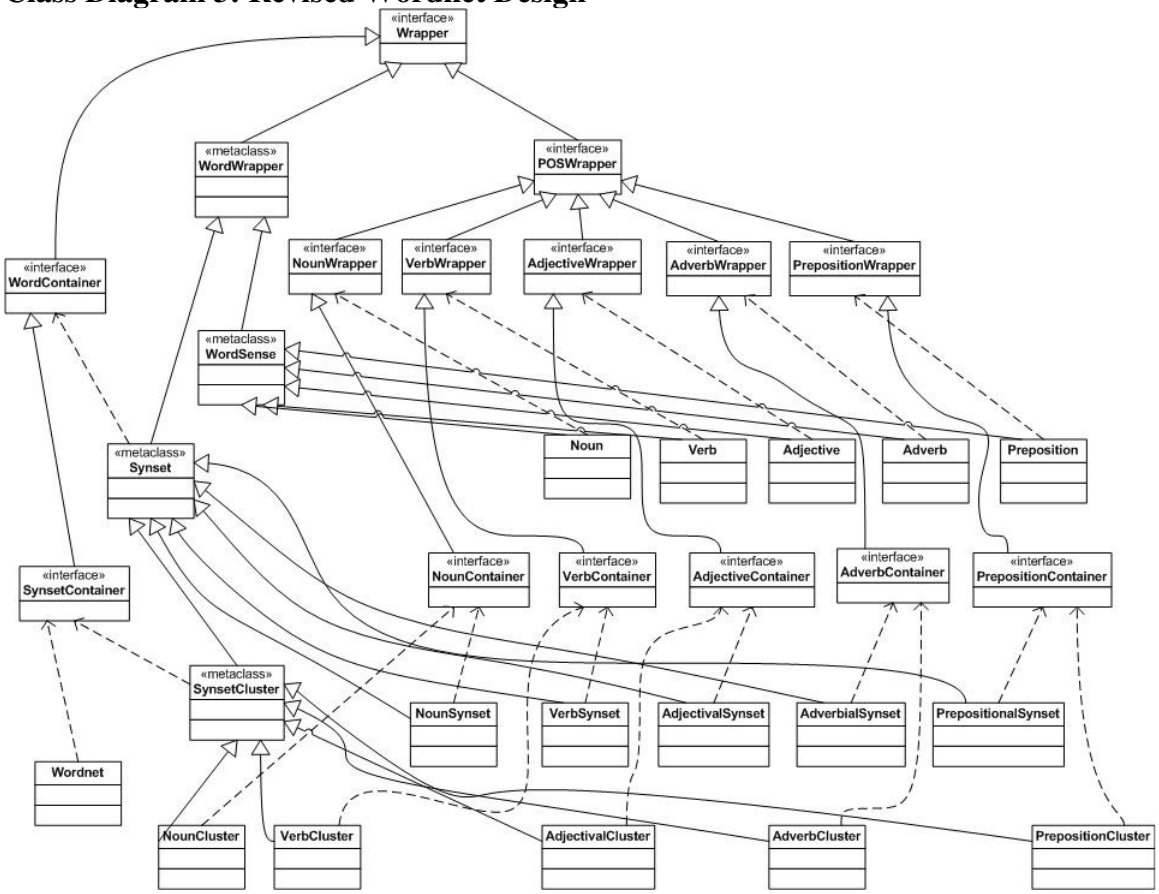
Class Diagram 1: Subclasses of Synset and WordSense



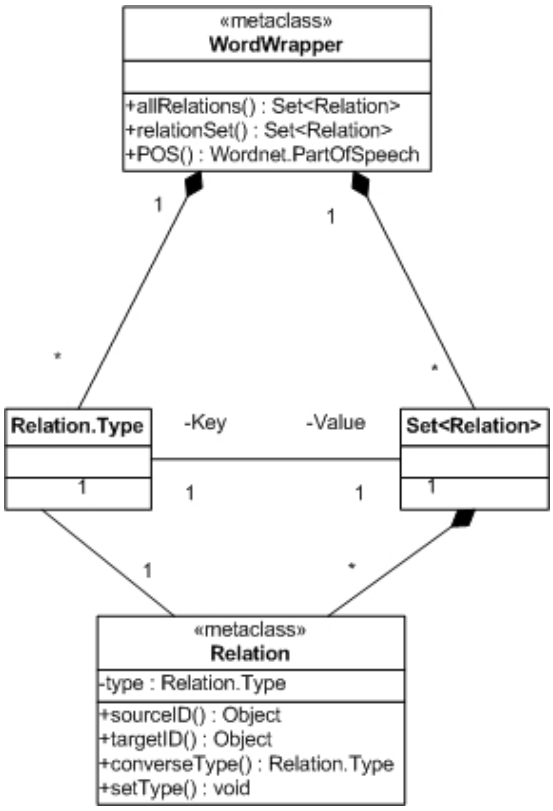
Class Diagram 2: Top Level Class Diagram of WordNet Model and Lexicon



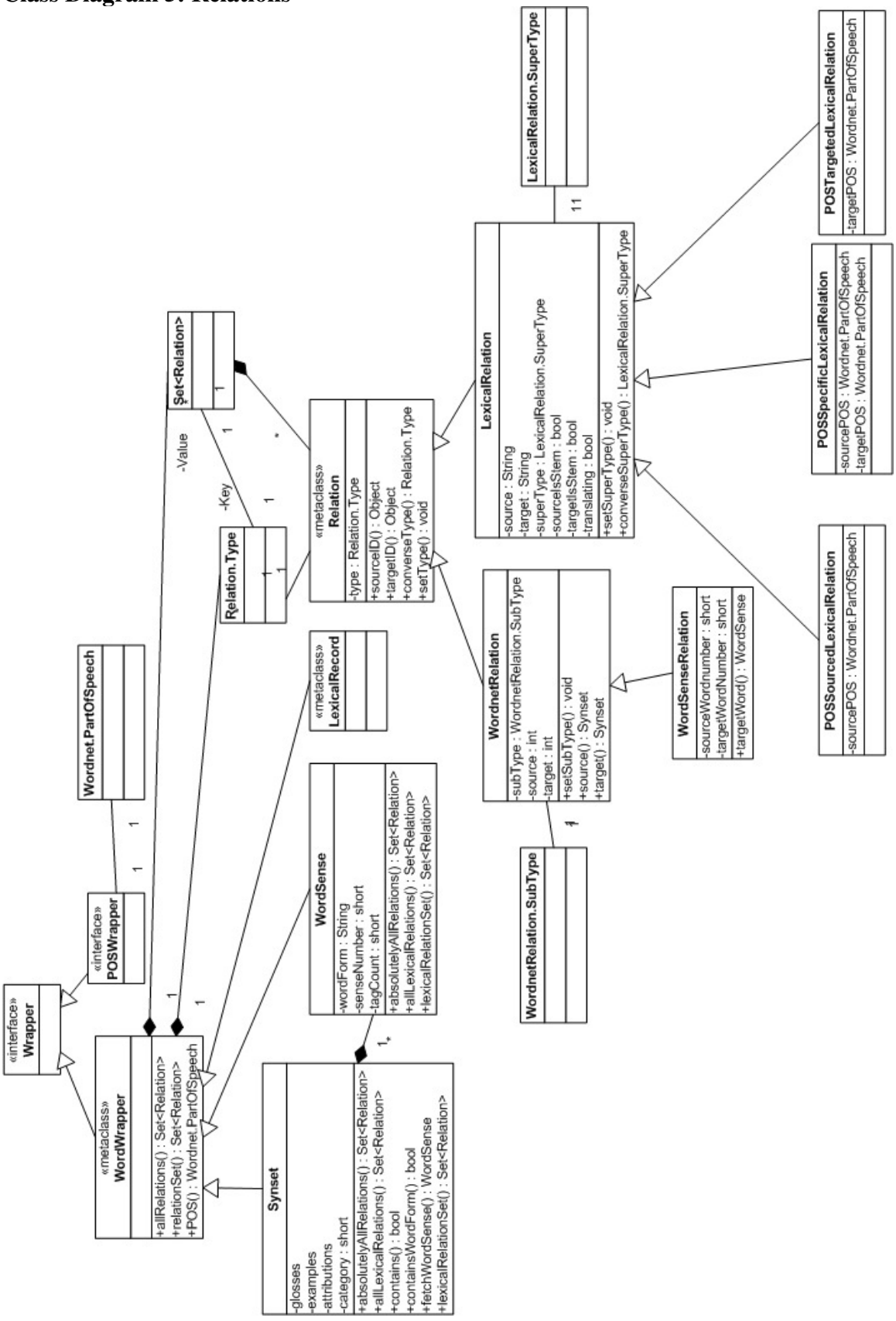
Class Diagram 3: Revised Wordnet Design



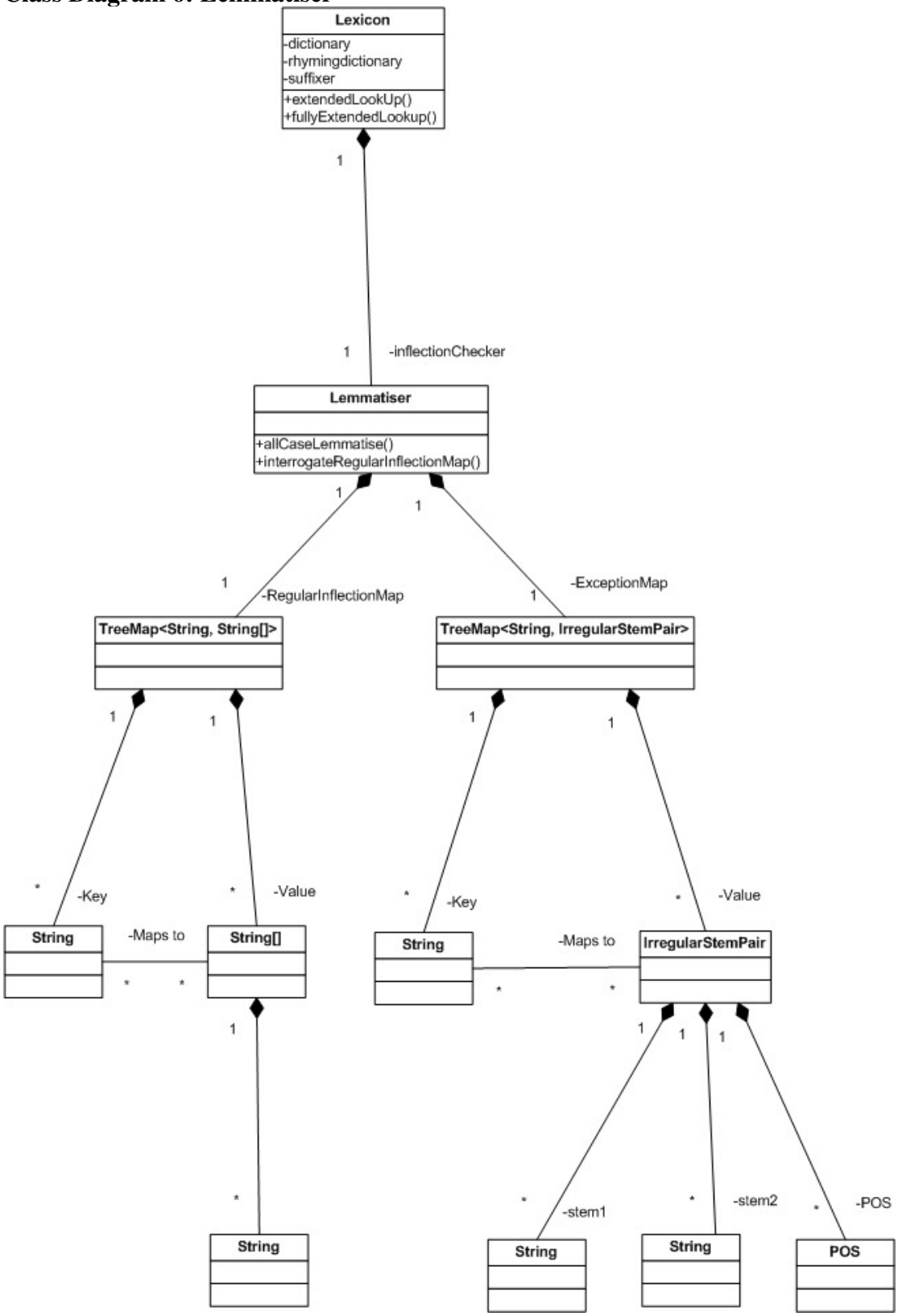
Class Diagram 4: WordWrapper Structure



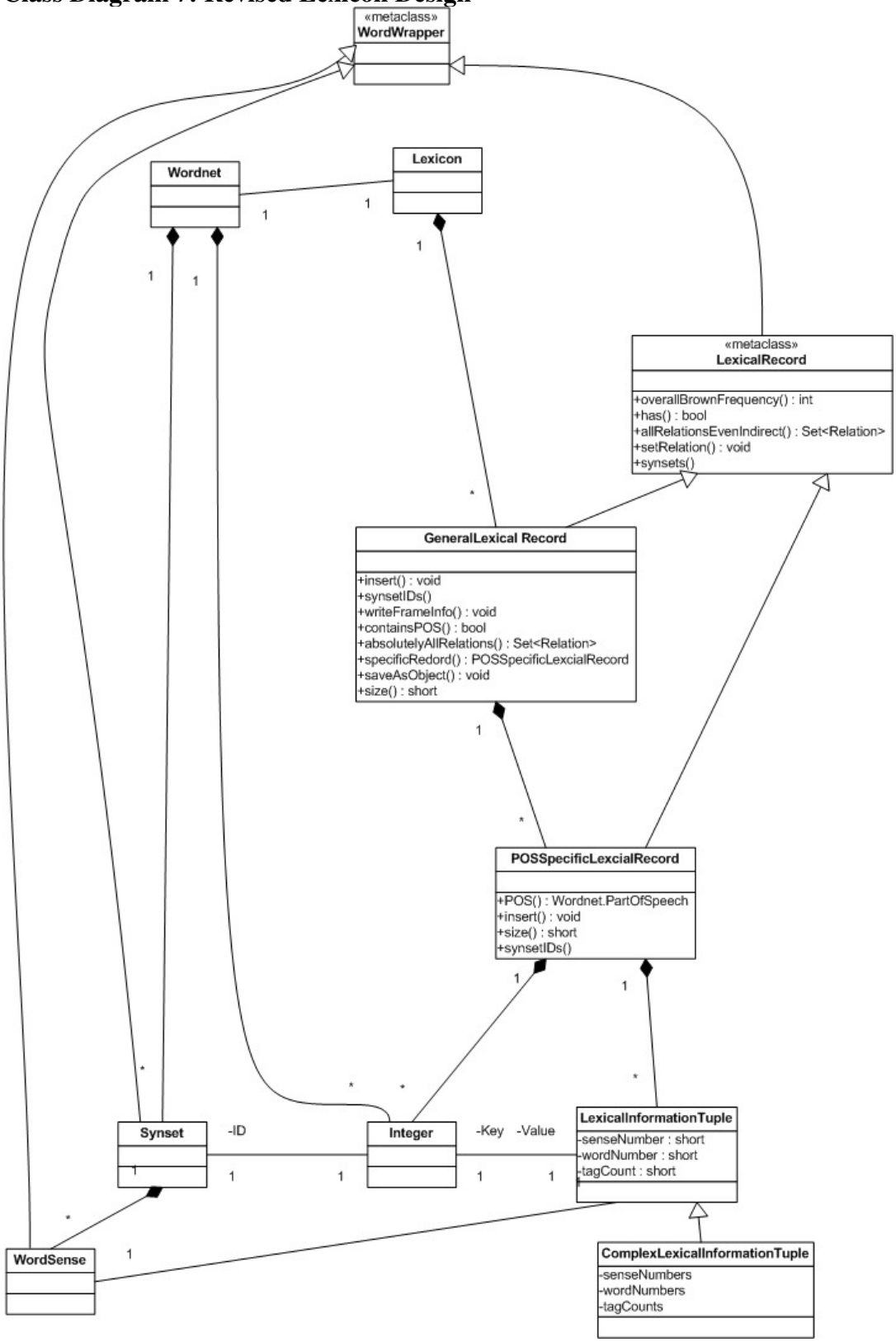
Class Diagram 5: Relations



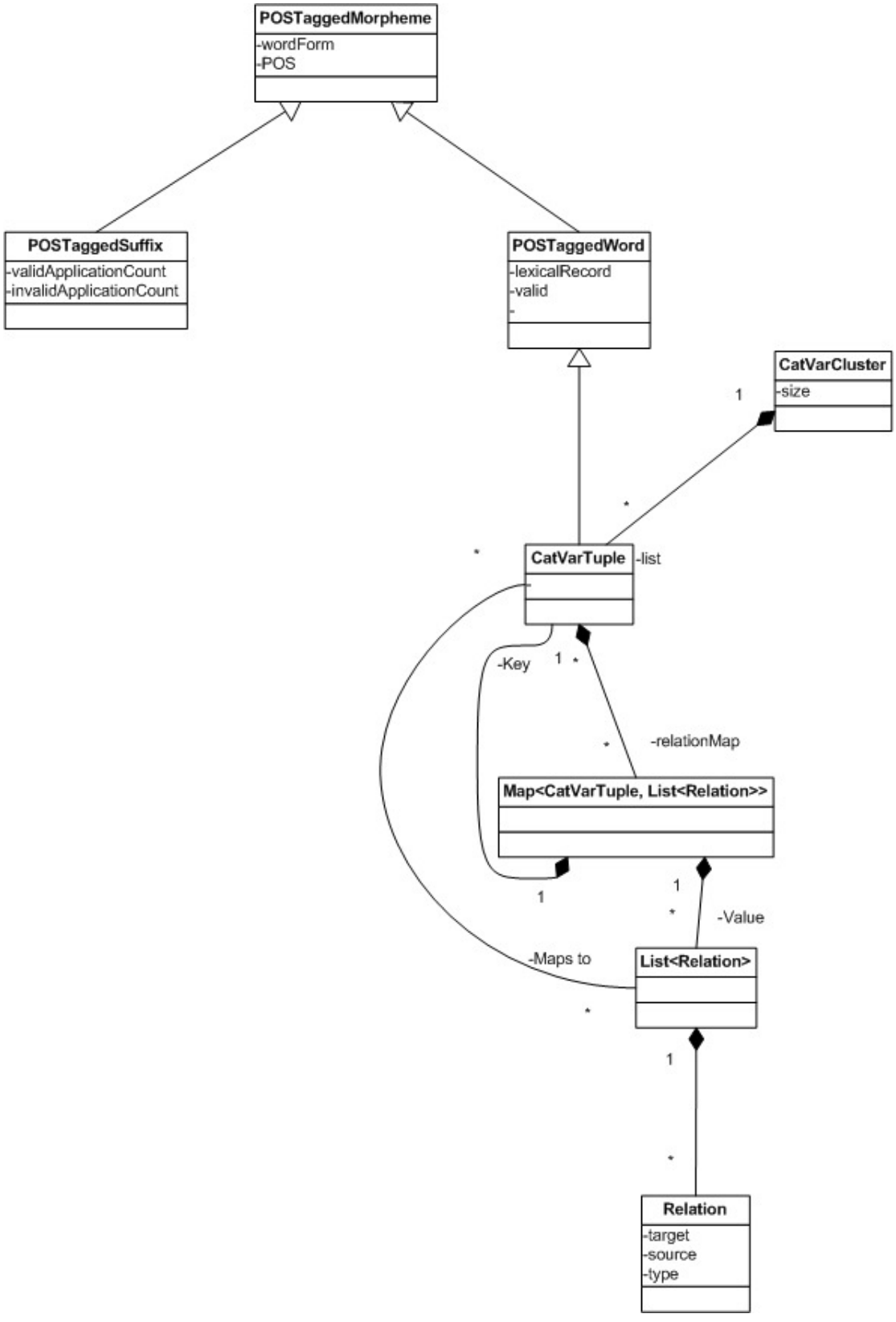
Class Diagram 6: Lemmatiser



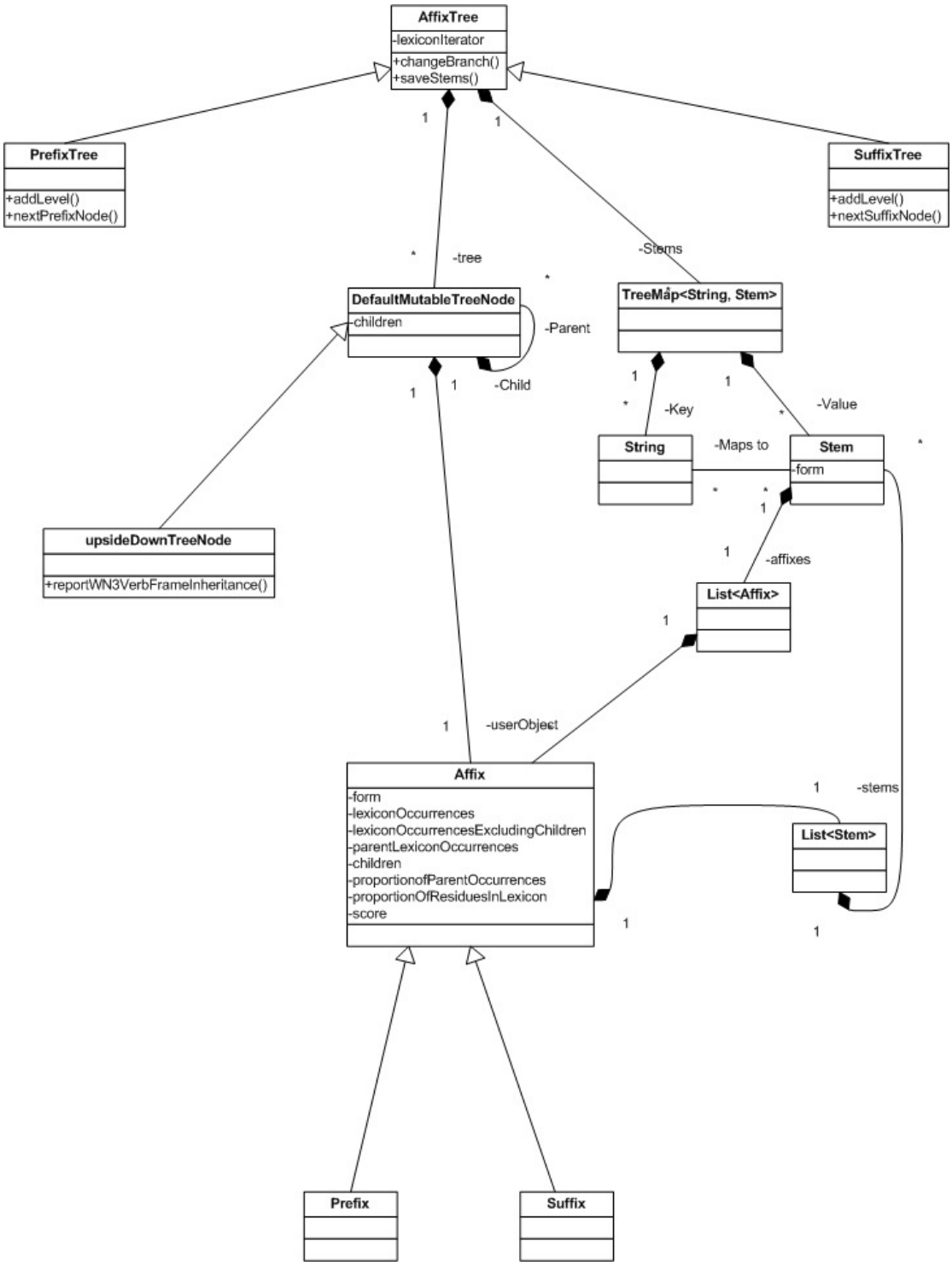
Class Diagram 7: Revised Lexicon Design



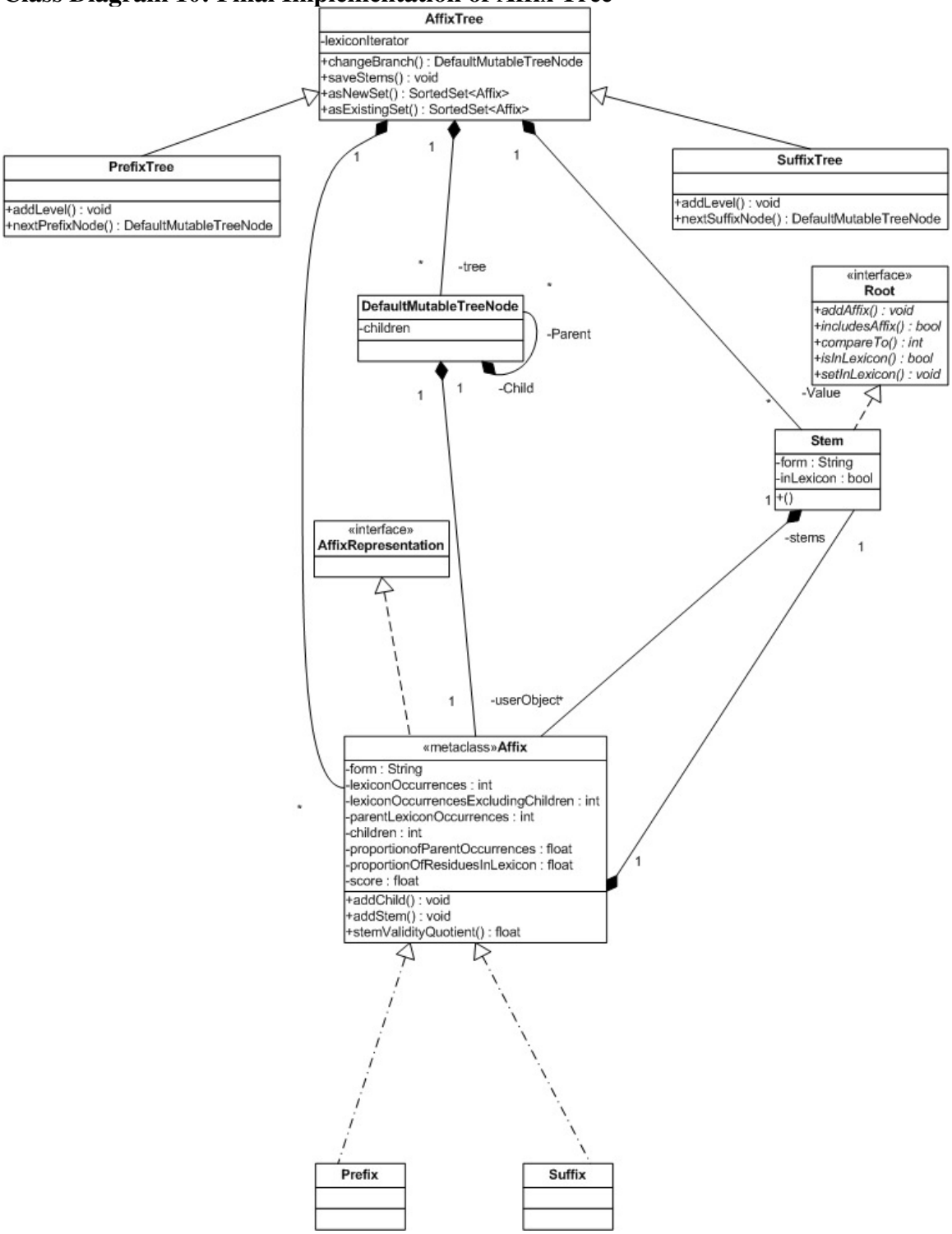
Class Diagram 8: Classes used to Represent CatVar Data and Morphological Rules



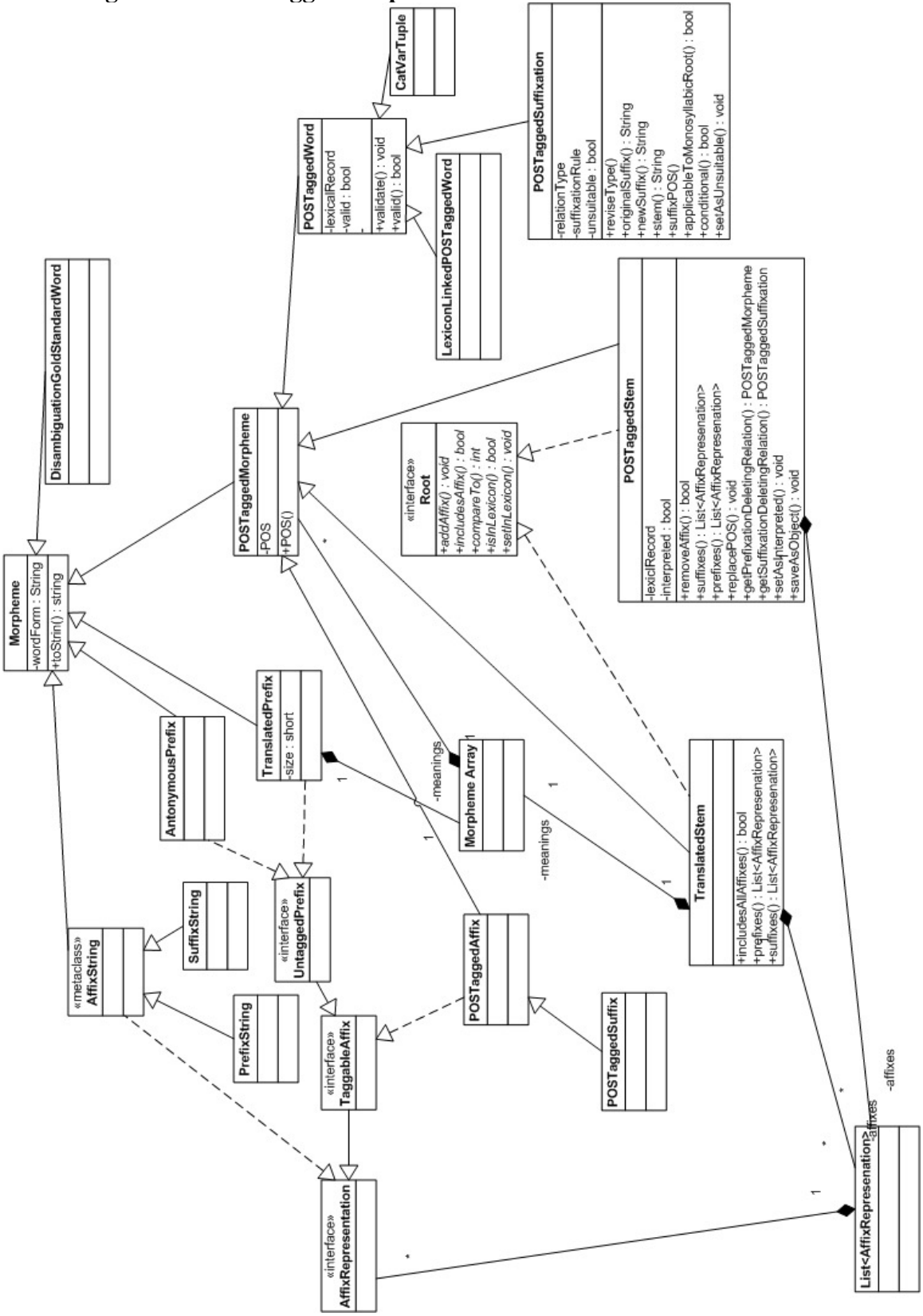
Class Diagram 9: Affix Tree



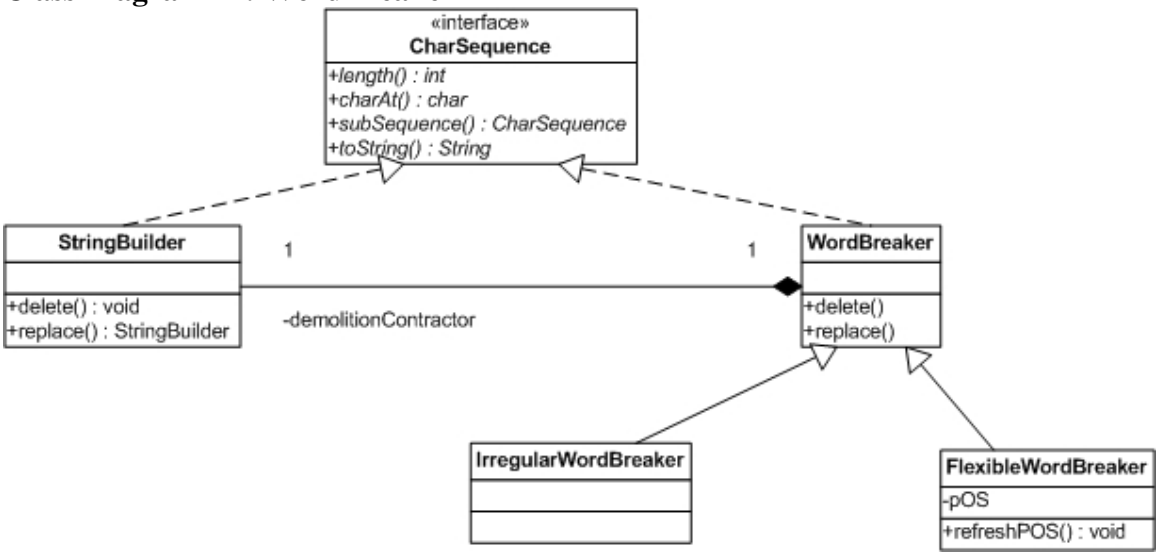
Class Diagram 10: Final Implementation of Affix Tree



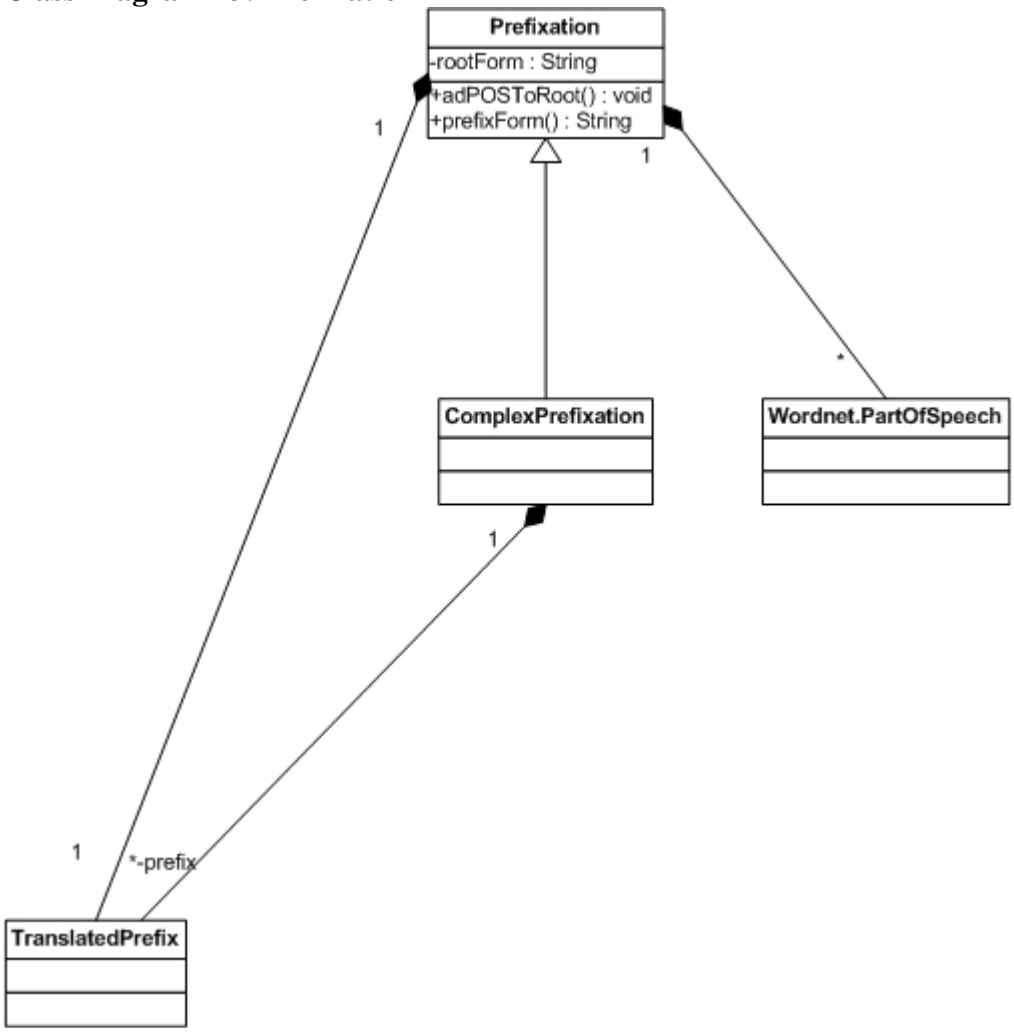
Class Diagram 11: POSTaggedMorpheme



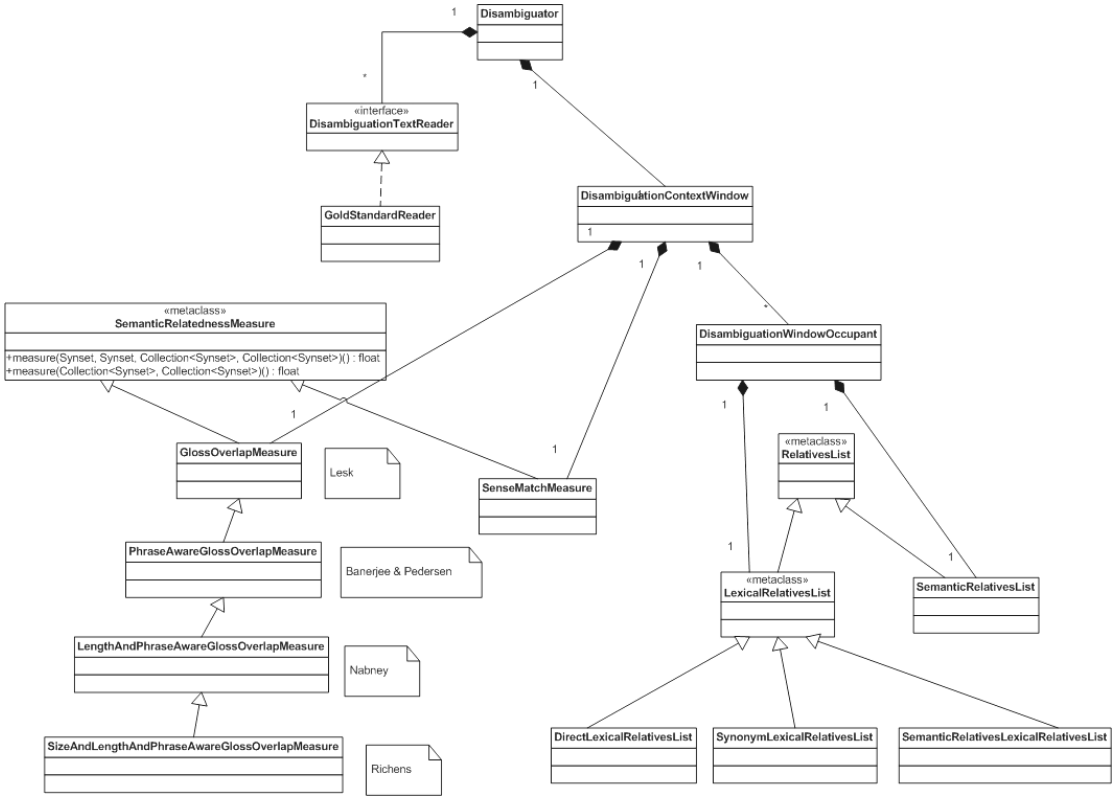
Class Diagram 12: WordBreaker



Class Diagram 13: Prefixation



Class Diagram 14: Disambiguator



Appendices

Appendix 1

Classes used to model WordNet and classes used in morphological analysis

For visualisation of the relationships between these classes in the most recent version, please refer to Class Diagrams 4, 5, 7, 10, 11 & 13.

```
public abstract class Affix
extends java.lang.Object
implements AffixRepresentation
```

Abstract class to represent an automatically discovered affix

```
public class Prefix
extends Affix
implements java.lang.Comparable
```

Class to represent an automatically discovered prefix

```
public class Suffix
extends Affix
implements java.lang.Comparable
```

Class to represent an automatically discovered suffix

```
public abstract class Affixer
extends java.lang.Object
```

Utility containing common functionality of `Prefixer` and `Suffixer`

```
public class Prefixer
extends Affixer
```

Class to handle the complexities of separating prefixes from their stems. Encapsulates 3 maps holding data about prefixes: the regular prefix translations `Map` maps from `String`s representing regular prefixes to `TranslatedPrefix`es; the irregular prefix translations `Map` maps from `String`s representing irregular prefixes to `TranslatedPrefix`es; the irregular prefixes `Map` maps from `String`s representing irregular prefix footprints to `List`s of `IrregularPrefixRecord`s.

```
public class Suffixer
extends Affixer
```

Utility class to handle the complexities of appending and removing suffixes. Encapsulates the morphological rules as mappings from `POSTaggedSuffix`es to `List`s of `MorphologicalRule`s of which the `POSTaggedSuffix` is the source, in the following maps: Unconditional morphological rules; Conditional morphological rules; Non-lexical morphological rules; Converse unconditional morphological rules; Converse conditional morphological rules; Converse non-lexical morphological rules; Non-lexical rules are default rules used in stem analysis. The conditional rules take

into account the irregular inflection data stored in the encapsulated exception map, which is the inverse of the exception map used by the lemmatiser and derived from the WordNet exception files. Converse rules are used for suffix stripping; the others are formulated for suffix application. The contents of both sets are the same except with source and target reversed and with the converse `Relation.Type`. A suffix stripping stoplist is encapsulated as mappings from `POSTaggedWord`s to `List`s of `POSTaggedWord`s, but is not initialised by the constructor.

```
public class AffixOrderer
extends java.lang.Object
implements java.util.Comparator<java.lang.String>,
java.io.Serializable
```

Comparator for comparing affixes represented as `String`s Imposes a primary ordering by affix length and a secondary lexicographic ordering.

```
abstract class AffixTree
extends java.lang.Object
```

Class to represent an affix tree rooted at an affix representing an empty string and encapsulating a `Set` of `Affix`es representing the contents of the tree ordered by a heuristic.

```
public class PrefixTree
extends AffixTree
```

Class to represent a prefix tree rooted at a prefix representing an empty string and encapsulating a `Set` of `Prefix`es representing the contents of the tree ordered by a heuristic.

```
public class SuffixTree
extends AffixTree
```

Class to represent a suffix tree rooted at a suffix representing an empty string and encapsulating a `Set` of `Suffix`es representing the contents of the tree ordered by a heuristic.

```
public class IrregularPrefixRecord
extends java.lang.Object
```

Class modelling an irregular prefix, encapsulating the corresponding footprint and `TranslatedPrefix` and the character `String`s to be deleted and inserted between the prefix and the stem when stripping the irregular prefix from a word. The `Set` of instances of words beginning with the prefix represented is also encapsulated.

```
public class IrregularStemPair
extends java.lang.Object
implements java.io.Serializable
```

Class encapsulating a maximum of 2 alternative stems and a `Wordnet.PartOfSpeech` for the stems of a word with irregular inflectional morphology across POS transformation. Most typically this Class encapsulates a single irregular verb

```
public final class Lemmatiser
extends java.lang.Object
implements java.io.Serializable
```

Utility for finding lemmas of inflected words. It encapsulates a regular inflection map and an exception map and a list of abbreviated inflections which are preceded by an apostrophe.

```
public class LexicalInformationTuple
extends java.lang.Object
implements java.io.Serializable, java.lang.Cloneable
```

Class to hold information in the `Lexicon` about a specific `WordSense`, comprising the sense number of the meaning of the word whose sense is represented, the word number of that word within the `Synset` which represents its meaning and a tag count, which represents the Brown Corpus frequency of the `WordSense`. The `LexicalInformationTuple` is held within a `POSSpecificLexicalRecord`.

```
public class ComplexLexicalInformationTuple
extends LexicalInformationTuple
```

An extension of `LexicalInformationTuple` representing multiple `WordSense`s. The fields are parallel arrays of the types of the fields in `LexicalInformationTuple`

```
public class LexicalPossibilityRecord
extends java.lang.Object
```

Class representing a word as a `String` and a `Set` of its possible POSes

```
public final class Lexicon
extends java.lang.Object
implements java.io.Serializable
```

Class implementing a lexicon based on WordNet encapsulating a main dictionary and optionally a rhyming dictionary, an atomic dictionary, a stem dictionary and an atomic stem dictionary. All these dictionaries, except the stem dictionary, map from `String`s representing words or stems. The main dictionary maps from a `String` corresponding to every word form or phrase in WordNet to the corresponding `GeneralLexicalRecord`. The rhyming dictionary maps from reversed word forms to `Set`s of their possible POSes. The atomic dictionary maps from words, which have not yet been broken down morphologically into their components, to sets of their possible POSes. The stem dictionary is a lexicographically ordered set of `POSTaggedStem`s from morphological analysis. The atomic stem dictionary maps from `String`s representing stems to `Set`s of their possible POSes.

```
public class Morpheme
extends java.lang.Object
implements java.lang.Comparable<Morpheme>, java.io.Serializable
```

Class representing a word or part of the word with no information except a `String` representing its orthography

```
public abstract class AffixString
extends Morpheme
implements AffixRepresentation
```

Class to represent an affix, holding no information except the `String` representing the form of the affix

```
public class PrefixString
extends AffixString
```

A representation of a prefix as a `String`

```
public class SuffixString
extends AffixString
```

A representation of a suffix as a `String`

```
public class AntonymousPrefix
extends Morpheme
implements UntaggedPrefix, java.io.Serializable
```

Class representing an antonymous prefix, holding no information except the `String` representing the form of the prefix

```
public class POSTaggedMorpheme
extends Morpheme
implements java.lang.Comparable<Morpheme>, java.io.Serializable
```

Holds a `String` representing a morpheme and the POS associated with it.

```
public abstract class POSTaggedAffix
extends POSTaggedMorpheme
implements TaggableAffix, java.io.Serializable
```

Class to represent an affix with a known form and POS

```
public class POSTaggedSuffix
extends POSTaggedAffix
implements java.io.Serializable
```

Holds a `String` representing a suffix and the POS associated with it.

```
public class POSTaggedStem
extends POSTaggedMorpheme
implements Root, java.io.Serializable
```

Class representing a stem with a known orthographic form and POS encapsulating lists of attested prefixes and suffixes and a `POSSpecificLexicalRecord`

```
public class POSTaggedWord
extends POSTaggedMorpheme
implements java.lang.Comparable<Morpheme>
```

Holds a `String` representing a word and the POS associated with it, along with a lexical record for it if it is in the lexicon as the specified POS.

```
public class LexiconLinkedPOSTaggedWord
extends POSTaggedWord
```

A version of `POSTaggedWord` which requires the corresponding `GeneralLexicalRecord` to be passed to its constructor

```
public class POSTaggedSuffixation
extends POSTaggedWord
implements java.lang.Comparable<Morpheme>
```

Class representing a word as a suffixation, encapsulating the `Relation.Type` which holds between it and its otherwise suffixed morphological derivative. The `MorphologicalRule` by which the suffixation is derived is also encapsulated, from which the new (current) suffix (if any) and the original suffix (of its derivative) can be extracted.

```
public class TranslatedStem
extends POSTaggedMorpheme
implements Root
```

Class representing a stem encapsulating `List`s of associated prefixes and suffixes as `AffixRepresentation`s and the stem's meanings as an array of `POSTaggedMorpheme`s

```
public class TranslatedPrefix
extends Morpheme
implements UntaggedPrefix, java.io.Serializable
```

Class representing a prefix and encapsulating its meanings as an array of `POSTaggedMorpheme`s

```
public class MorphologicalAnalyser
extends java.lang.Object
```

Class for performing morphological analysis tasks on data from the `Lexicon`, encapsulating (references to) the `NaturalLanguageProcessor`, `Lexicon`, `Prefixer`, `Suffixer`, `Wordnet`, `Lemmatiser` and `Lexicon` fields dictionary, `rhymingDictionary`, `atomicDictionary`, `stemDictionary` and `atomicStemDictionary` along with a constant `String` array of antonymous prefixes namely "un", "in", "imb", "ign", "ill", "imm", "imp", "irr", "dis", "de", "counter", "contra", "contr", "non", "anti", "ant", "an", "a"

```
public class MorphologicalRule
extends java.lang.Object
implements java.lang.Comparable<MorphologicalRule>
```

Class to model a morphological rule. It encapsulates 2 `POSTaggedSuffix`es as the source and target of the rule. The rule represents a transformation from the source to

the target. The `Relation.Type` of the relation from the source to the target is also encapsulated. A Boolean field defines whether the rule is conditional, meaning that it can be overridden by irregular participle formation or ADJECTIVE/ADVERB comparison Another Boolean field specifies whether the rule is applicable to a transformation between a derivative and a root when the root is monosyllabic, irrespective of whether the root is the source or the target.

```
public class MorphoSemanticWordnetBuilder
extends java.lang.Object
```

Utility for specifying and processing morphological analyses conducted by the `MorphologicalAnalyser`.

```
public class MutableCollection
extends java.lang.Object
```

Houses a `Collection` which can be either a `List` or a `Set` at different times depending on the required functionality. It is used to store `VerbFrame`s.

```
public final class NaturalLanguageProcessor
extends java.lang.Object
```

Top level class encapsulating the entire model. It encapsulates the `Wordnet`, `Lexicon`, `Lemmatiser`, `Prefixer` and `Secator` and optionally a `MutableCollection` of `VerbFrame`s.

```
public class OptimalHeuristic
extends java.lang.Object
implements java.util.Comparator<Affix>
```

Comparator to compare 2 `Affix`es according to the optimal heuristic

$$\frac{f_c^2 q_s}{f_p}$$

where f_c = affix frequency, f_p = parent frequency and q_s = stem validity quotient. A secondary ordering is imposed by affix frequency and a tertiary ordering by orthographic form.

```
public class Prefixation
extends java.lang.Object
```

Class to represent a word comprising a prefix and a stem, encapsulating a `String` a `Set` of possible POSes representing the stem and a `TranslatedPrefix` representing the prefix

```
public class ComplexPrefixation
extends Prefixation
```

An extension of `Prefixation` allowing multiple `TranslatedPrefix`es

```
public class PrefixLengthComparator
extends java.lang.Object
implements java.util.Comparator<Morpheme>
```

Comparator for comparing prefixes as `Morpheme`s. Prioritises the longest prefixes.

```
public class PTMComparator
extends java.lang.Object
implements java.util.Comparator<POSTaggedMorpheme>,
java.io.Serializable
```

Comparator for comparing `POSTaggedMorpheme`s. Imposes a primary lexicographic ordering and a secondary ordering by POS.

```
public class PTSuffixationComparator
extends java.lang.Object
implements java.util.Comparator<POSTaggedSuffixation>,
java.io.Serializable
```

Comparator for comparing `POSTaggedSuffixation`s. Imposes a primary ordering by `Relation.Type`, secondary lexicographic ordering and tertiary ordering by POS.

```
public class PTSuffixationFrequencyComparator
extends java.lang.Object
implements java.util.Comparator<POSTaggedSuffixation>,
java.io.Serializable
```

Comparator for comparing `POSTaggedSuffixation`s. Imposes an ordering by Brown Corpus Frequency.

```
public class PTSuffixComparator
extends java.lang.Object
implements java.util.Comparator<POSTaggedSuffix>,
java.io.Serializable
```

Comparator for comparing `POSTaggedSuffix`es. Imposes a primary ordering by word length and a secondary lexicographic ordering.

```
public abstract class Relation
extends java.lang.Object
implements java.io.Serializable
```

Class representing a relationship between from one Object (the source) to another Object (the target), both of which have a corresponding `WordWrapper` (`Synset`, `WordSense` or `LexicalRecord`). Every `Relation` has a `Relation.Type` which is one of the following: {HYPERNYM, HYPONYM, ENTAILMENT, COUNTER_ENTAILMENT, CAUSE, EFFECT, INSTANCE, INSTANTIATED, SIMILAR, CLUSTERHEAD, MEMBER_MERONYM, MEMBER_HOLONYM, SUBSTANCE_MERONYM, SUBSTANCE_HOLONYM, PART_MERONYM, PART_HOLONYM, ATTRIBUTE, ATTRIBUTE_VALUE, CLASS_MEMBER, MEMBER_CLASS, SEE_ALSO, SEEN_ALREADY, PARTICIPLE, VERB_SOURCE, PERTAINYM, PERTAINER, ROOT, DERIVATIVE, ANTONYM_OF_ATTRIBUTE_VALUE, ATTRIBUTE_OF_ANTONYM,

ANTONYM_OF_PARTICIPLE, VERBSOURCE_OF_ANTONYM, GERUND, VERBSOURCE_OF_GERUND, MEASUREDBY, MEASURING, PATIENT, AFFECTING, ABLE, POTENTIAL, QUALIFIED, QUALIFYING, RESEMBLING, RESEMBLEDBY, DEMONSTRATE, DEMONSTRATION, SUBJECT, ROLE, POSSESSION_OF_ATTRIBUTE, POSSESSOR_OF_ATTRIBUTE, SUBJECT_OF_VERBSOURCE_OF_GERUND, GERUND_OF_ROLE, BELIEVE_PRACTICE, OBJECT_OF_BELIEF_PRACTICE, GERUND_OF_BELIEVE_PRACTICE, OBJECT_OF_BELIEF_PRACTICE_OF_VERBSOURCE_OF_GERUND, GERUND_OF_BELIEVE_PRACTICE_PERTAINYM, PERTAINER_TO_OBJECT_OF_BELIEF_PRACTICE_OF_VERBSOURCE_OF_GERUND, SUBJECT_OF_BELIEVE_PRACTICE, OBJECT_OF_BELIEF_PRACTICE_OF_ROLE, SUBJECT_OF_BELIEVE_PRACTICE_PERTAINYM, PERTAINER_TO_OBJECT_OF_BELIEF_PRACTICE_OF_ROLE, SINGULAR, PLURAL, MASCULINE, FEMININE, DESTINATION, DIRECTION, COMPARISON, ADJECTIVE_SOURCE, HOME, INHABITANT, FULLSIZE, DIMINUTIVE, REPEATED, REPETITION, AFFECTED_ORGAN, DISEASE, ABILITY, POTENTIALITY, ANTONYM, VERB_GROUP_POINTER, DERIV, NEARSYNONYM, SYNONYM}. Every `Relation` has a converse, where the source and target are reversed. The `Relation.Type` of the converse `Relation` must be the converse type of the first `Relation`'s `Relation.Type`. `Relation.Type`s in the above list are in pairs, each of which is the converse of the other, except for the last 5, where the converse type is the same type. `Relation.Type` pairs may be added to the list, but the five types which are their own converses are invariant in number and must remain at the end of the list.

```
public class LexicalRelation
extends Relation
```

Class representing a morphological relationship between two morphemes (either words or stems) represented as `String`s, the source, in whose corresponding `LexicalRecord` this `LexicalRelation` is encoded, and a target. The status of the source and target as a word or a stem are held in Boolean fields. Another Boolean field specifies whether either source or target (never both) is a translation of a stem or prefix. Every `LexicalRelation` has a `LexicalRelation.SuperType` which is either DERIVATIVE (if the target is derived from the source), or ROOT (if the source is derived from the target). The `LexicalRelation.SuperType` must be consistent with the inherited `Relation.Type`. If the `LexicalRelation.SuperType` is ROOT then the `Relation.Type` must be the first of a pair in the list of `Relation.Type`s listed under `Relation` above or one of the 5 types which are their own converses; if the `LexicalRelation.SuperType` is DERIVATIVE then the `Relation.Type` must be the second of a pair in the list of `Relation.Type`s or one of the 5 types which are their own converses.

```
public class POSSourcedLexicalRelation
extends LexicalRelation
```

Class representing a morphological relation between two words of which the POS of the source is specified

```
public class POSSpecificLexicalRelation
extends LexicalRelation
```

Class representing a morphological relation between two words both of whose POSes are specified

```
public class POSTargetedLexicalRelation
extends LexicalRelation
```

Class representing a morphological relation between two words of which the POS of the target is specified

```
public class WordnetRelation
extends Relation
```

Class representing a semantic relationship between two `Synset`s represented by integers which are `Synset` identifiers, the source, where this `LexicalRelation` is encoded, and a target. A `WordnetRelation` may have a `subType`.

```
public class WordSenseRelation
extends WordnetRelation
implements java.io.Serializable
```

Class representing a morphosemantic relationship between two `WordSense`s, whose `Synset` identifiers are represented by integers and whose word numbers within those `Synset`s are also specified.

```
public class Secator
extends java.lang.Object
```

Utility for pruning the `Wordnet`.

```
public class Stem
extends java.lang.Object
implements Root
```

Class to represent the residue of an affixation after removal of the affix during automatic affix discovery

```
abstract class VerbFrame
extends java.lang.Object
implements MutableCollectionMember, java.io.Serializable
```

Defines common functionality of WordNet and parse-generated verb frames with respect to valency (number of arguments) and verb frame inheritance.

```
public class WordNetVerbFrame
extends VerbFrame
implements java.io.Serializable,
java.lang.Comparable<WordNetVerbFrame>
```

Class representing any of the 35 WordNet verb frames.

```
public class WordBreaker
extends java.lang.Object
implements java.lang.CharSequence
```

Utility Class which ideally would expand `StringBuilder`, but as `StringBuilder` is **`final`**, it implements `CharSequence`, as does `StringBuilder` and contains a `StringBuilder` field. It encapsulates references to the `Prefixer`, `Suffixer`, `Lexicon`, `Wordnet` and `Lemmatiser`. The embedded `StringBuilder` contains a word, which is reduced to its stem by the `WordBreaker`'s delete method which removes an affix.

```
public class FlexibleWordBreaker
extends WordBreaker
```

Utility Class extending `WordBreaker` and encapsulating a `Wordnet.PartOfSpeech`, for representing a stem during stem analysis. The stem is reduced to a shorter stem by the `FlexibleWordBreaker`'s delete method.

```
public class IrregularWordBreaker
extends WordBreaker
```

Extension of `WordBreaker` to encapsulate an irregular prefixation. Its delete method removes the irregular prefix leaving the stem.

```
public final class Wordnet
extends java.lang.Object
implements java.io.Serializable, SynsetContainer
```

Class modelling Princeton WordNet. The `Synset`s are held in a map from which they are retrieved using the `Synset` ID as a key. A record is kept of the next available `Synset` ID for each POS.

```
public abstract class WordWrapper
extends java.lang.Object
implements Wrapper, java.io.Serializable
```

Abstract Class to hold the common functionality of `Synset`, `WordSense` and `LexicalRecord`, namely a `Map<WordnetBuilder.Relation.Type, Set<Relation>>`, in which the `Relation.Type`s permitted for the particular subclass map to the `Relation`s whose source is the `Synset` identifier, or the `Synset` identifier of the `Synset` which contains the `WordSense` or the word which maps to the `LexicalRecord` in the main dictionary of the `Lexicon`.

```
public abstract class LexicalRecord
extends WordWrapper
implements java.io.Serializable
```

Abstract class encapsulating the common fields and methods of a `GeneralLexicalRecord` or `POSSpecificLexicalRecord` held in the main dictionary of the `Lexicon`. Holds `LexicalRelation`s targeted on words or stems. Normally held in the main dictionary of the `Lexicon`, but can also be encapsulated in a `POSTaggedStem` in the stem dictionary.

```
public class GeneralLexicalRecord
extends LexicalRecord
implements java.io.Serializable, java.lang.Cloneable
```

Class encapsulating the information held about a word in the main dictionary of the `Lexicon`. The information maps from each possible `Wordnet.PartOfSpeech` of the word to which this `GeneralLexicalRecord` refers to the corresponding `POSSpecificLexicalRecord`. Holds `LexicalRelation`s targeted on words or stems.

```
public abstract class POSSpecificLexicalRecord
extends LexicalRecord
implements java.io.Serializable
```

Class to encapsulate the information held in the `Lexicon` about a word as a wordform with a specified POS. The information is held as mappings from Integers representing `Synset` IDs to `LexicalInformationTuple`s. Holds `LexicalRelation`s targeted on words or stems. Can be encapsulated in a `POSTaggedStem` in the stem dictionary, but without any `LexicalInformationTuple`s.

```
public abstract class Synset
extends WordWrapper
implements java.io.Serializable, WordContainer
```

Represents a synset as in WordNet. It holds a semantic category number and a list of `WordSense`s. The WordNet gloss is subdivided into a set of `String`s representing the actual glosses and 2 co-indexed lists of `String`s representing the, examples and their attributions.

```
public abstract class WordSense
extends WordWrapper
implements java.io.Serializable, java.lang.Cloneable
```

Represents a word sense as in WordNet, which is the intersection of one word and one meaning. It hold the word form, which may be a multiword expression and the sense number of the particular senses of the word. It also holds a tag count which represents its frequency in the sense-tagged Brown corpus. The WordNet sense key is stored as its separate components according to the WordNet documentation.

Appendix 2

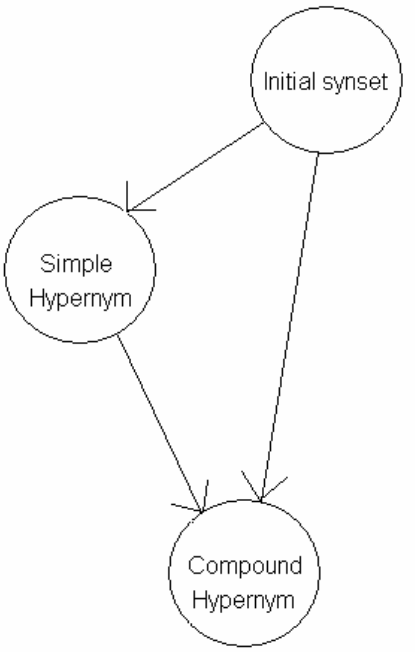
WordNet verb frames

```
1     Something ----s
2     Somebody ----s
3     It is ----ing
4     Something is ----ing PP
5     Something ----s something Adjective/Noun
6     Something ----s Adjective/Noun
7     Somebody ----s Adjective
8     Somebody ----s something
9     Somebody ----s somebody
10    Something ----s somebody
11    Something ----s something
12    Something ----s to somebody
13    Somebody ----s on something
14    Somebody ----s somebody something
15    Somebody ----s something to somebody
16    Somebody ----s something from somebody
17    Somebody ----s somebody with something
18    Somebody ----s somebody of something
19    Somebody ----s something on somebody
20    Somebody ----s somebody PP
21    Somebody ----s something PP
22    Somebody ----s PP
23    Somebody's (body part) ----s
24    Somebody ----s somebody to INFINITIVE
25    Somebody ----s somebody INFINITIVE
26    Somebody ----s that CLAUSE
27    Somebody ----s to somebody
28    Somebody ----s to INFINITIVE
29    Somebody ----s whether INFINITIVE
30    Somebody ----s somebody into V-ing something
31    Somebody ----s something with something
32    Somebody ----s INFINITIVE
33    Somebody ----s VERB-ing
34    It ----s that CLAUSE
35    Something ----s INFINITIVE
```

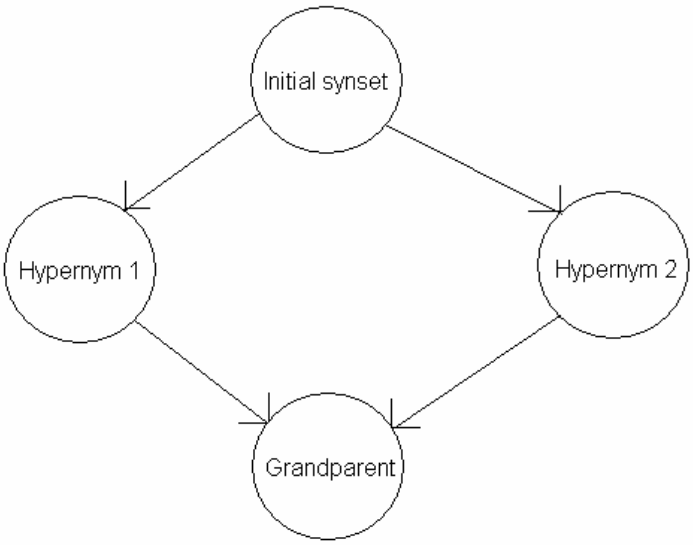
Appendix 3

Ring topologies

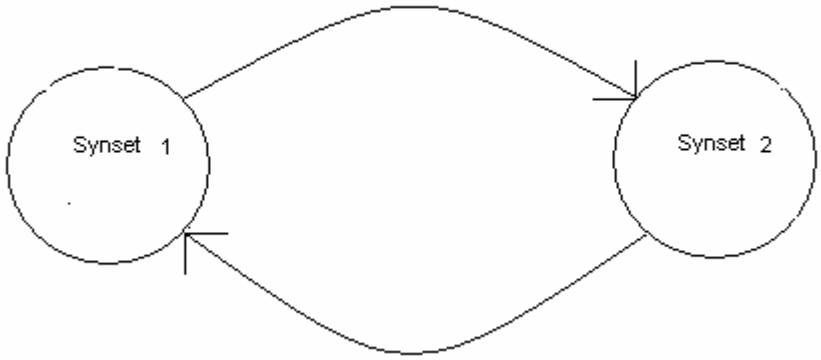
(a) Asymmetric topology



(b) Symmetric topology



(c) Cycle topology



Appendix 4

WordNet verb categories (*after Liu et al., 2004*)

29 Body
30 Change
31 Cognition
32 Communication
33 Competition
34 Consumption
35 Contact
36 Creation
37 Emotion
38 Motion
39 Perception
40 Possession
41 Social
42 Stative
43 Weather

Appendix 5

Valency and frame inheritance

Abbreviations in the table:

Fr. Frame
Val. Valency
Gov. Governed
Re-arr. Rearranged
V Verb
n. Noun
adj. Adjective
TH Theme
AG Agent
PAT Patient

INSTR Instrument
CL Clause
Pred. Predicate
Inf. Infinitive
Part. Active participle
Subj. Subject
D. Obj. Direct object
I. Obj. Indirect object
Gen. Genitive
Abl. Ablative
Obliq. Oblique case

Fr.	Condensed WordNet representation	Val.	Inherits	Adds	As	Gov. by	Re-arr.	As	Gov. by
3	It is ..ing	0							
1	Something ..s	1	3	TH	Subj.				
2	Somebody ..s	1	3	AG	Subj.				
34	It ..s that CLAUSE	1	3	CL	Pred.	*that*			
4	Something is ..ing PP	2	1	?	Obliq.	?			
6	Something ..s adj./n.	2	1	adj./n.	Pred.				
7	Somebody ..s adj.	2	2	adj.	Pred.				
8	Somebody ..s something	2	2	TH	D. Obj.				
9	Somebody ..s somebody	2	2	PAT	D. Obj.				
10	Something ..s somebody	2	1	PAT	D. Obj.				
11	Something ..s something	2	1	TH	D. Obj.				
12	Something ..s to somebody	2	1	PAT	I. Obj.	*to*			
13	Somebody ..s on something	2	2	?	Obliq.	*on*			
22	Somebody ..s PP	2	2	?	Obliq.	?			
23	Somebody's (body part) ..s	1.5	8				AG TH	Gen. Subj.	
26	Somebody ..s that CLAUSE	2	2,34				CL	D. Obj.	*that*
27	Somebody ..s to somebody	2	2	PAT	I. Obj.	*to*			
28	Somebody ..s to INFINITIVE	2	2	V	Inf.	*to*			
29	Somebody ..s whether INFINITIVE	2	2	V	Inf.	*whether to*			

Fr.	Condensed WordNet representation	Val.	Inherits	Adds	As	Gov. by	Re-arr.	As	Gov. by
32	Somebody ..s INFINITIVE	2	2	V	Inf.				
33	Somebody ..s Ving	2	2	V	Part.				
35	Something ..s INFINITIVE	2	1	V	Inf.				
5	Something ..s something adj./n.	3	6,11				adj./n.	Result	
14	Somebody ..s somebody something	3	8,9				PAT	I. Obj.	
15	Somebody ..s something to somebody	3	8,9				PAT	I. Obj.	*to*
16	Somebody ..s something from somebody	3	8,9				PAT	Abl.	*from*
17	Somebody ..s somebody with something	3	8,9				INSTR	Obliq.	*with*
18	Somebody ..s somebody of something	3	8,9				TH	Obliq.	*of*
19	Somebody ..s something on somebody	3	8,9				PAT	Obliq.	*on*
20	Somebody ..s somebody PP	3	9,22						
21	Somebody ..s something PP	3	8,22						
24	Somebody ..s somebody to INFINITIVE	3	9,28						
25	Somebody ..s somebody INFINITIVE	3	9,32						
31	Somebody ..s something with something	3	8	INSTR	Obliq.	*with*			
30	Somebody ..s somebody into Ving something	4	8,9,33				V	Part.	*into*

Appendix 6

Valid inheritance by tightening selectional restrictions
(*for abbreviations used, see Appendix 5*)

Fr.	Condensed WordNet representation	Val.	Inherits	Condensed WordNet representation	Val.
2	Somebody ..s	1	1	Something ..s	1
23	Somebody's (body part) ..s	1.5			
7	Somebody ..s adj.	2	6	Something ..s adj./n.	2
8	Somebody ..s something	2	11	Something ..s something	2
10	Something ..s somebody	2			
9	Somebody ..s somebody	2	8	Somebody ..s something	2
			10	Something ..s somebody	2
12	Something ..s to somebody	2	4	Something is ..ing PP	2
22	Somebody ..s PP	2			
13	Somebody ..s on something	2	22	Somebody ..s PP	2
27	Somebody ..s to somebody	2	12	Something ..s to somebody	2
32	Somebody ..s INFINITIVE	2	35	Something ..s INFINITIVE	2
28	Somebody ..s to INFINITIVE	2			
15	Somebody ..s something to somebody	3	21	Somebody ..s something PP	3
16	Somebody ..s something from somebody	3			
19	Somebody ..s something on somebody	3			
20	Somebody ..s somebody PP	3			
17	Somebody ..s somebody with something	3	31	Somebody ..s something with something	3

Appendix 7

Evaluation of hypernym / troponym relations between verbal synsets in sample violating the relaxed rules for frame inheritance

Evaluation of relation	Instances
OK	22
Indirect	5
Reversed	2
None	4
Indeterminate	1
Hypernym is cause of troponym	1
Hypernym is cause of true hypernym	1
True hypernym is cause of encoded hypernym	1
Troponym inherits causative sense	1
Troponym inherits inchoative sense	1
Troponym inherits intransitive frameset	1
Intransitive frameset inherits intransitive sense	1
1 frameset inherits from hypernym	1
Troponym inherits 1 frameset	2
Hypernym needs to be split between true hypernym and hypernym of hypernym	1
Troponym entails passive of hypernym	1
Other syntactic alternation	2
28, 35 not inherited	1
28 not inherited	1
Troponym incorporates preposition	1
Hypernym incorporates preposition	1
Troponym incorporates complement	1
TOTAL	53

Appendix 8

CatVar cluster members unrelated to headword

Headword		Unrelated cluster members	
Bai	**NOUN**		
		bay	NOUN
		bay	VERB
		bay	ADJECTIVE
chilli	**NOUN**		
		chilly	ADJECTIVE
		chilliness	NOUN
chopin	**NOUN**		
		chopine	NOUN
compass	**NOUN**		
		compassion	NOUN
		compassionate	VERB

Headword		Unrelated cluster members	
		compassionate	ADJECTIVE
		compassionately	ADVERB
		compassionateness	NOUN
curse	**NOUN**		
		cursor	NOUN
fall	**NOUN**		
		fallal	NOUN
illegal	**ADJECTIVE**		
		illegible	ADJECTIVE
		illegibly	ADVERB
		illegibility	NOUN
mate	**VERB**		
		mater	NOUN
more	**NOUN**		
		mores	NOUN
mull	**NOUN**		
		mullion	NOUN
		mullioned	ADJECTIVE
orang	**NOUN**		
		orange	NOUN
		orange	ADJECTIVE
		orangeness	NOUN
overlie	**VERB**		
		overly	ADVERB
pally	**ADJECTIVE**		
		palliative	NOUN
		palliative	ADJECTIVE
revere	**NOUN**		
		revere	VERB
		revered	ADJECTIVE
		reverence	NOUN
		reverence	VERB
		reverent	ADJECTIVE
		reverently	ADVERB
		reverential	ADJECTIVE
		reverentially	ADVERB
spin	**NOUN**		
		spinal	NOUN
		spinal	ADJECTIVE
		spinally	ADVERB
squash	**NOUN**		
		squash	VERB
		squashed	ADJECTIVE

Headword		Unrelated cluster members	
still	**NOUN**		
		still	VERB
		still	ADJECTIVE
		still	ADVERB
		stillness	NOUN
stud	**NOUN**		
		student	NOUN
tie	**NOUN**		
		tier	NOUN
unanimity	**NOUN**		
		unanimated	ADJECTIVE
underseal	**NOUN**		
		undersize	ADJECTIVE
		undersized	ADJECTIVE
vie	**VERB**		
		vial	NOUN

Appendix 9

Morphological rules formulated.

Rules wholly or partly in italics refer to languages other than English. Some of these rules have been implemented without reference to those languages. Rules wholly in italics have not been implemented.

[Rules which overgenerated from the CatVar headwords and were excluded from the restricted ruleset are enclosed within square brackets.]

General suffixation rules

NB For these rules "y" is treated as a vowel

To add a suffix beginning with a vowel to a stem:

> if the stem ends in a single consonant, excluding "w" and "x", preceded by a single vowel (or vowel preceded by "qu"), unless the stem ends in "er", "or" or "om", if the stem is monosyllabic, the consonant is doubled before adding the suffix, otherwise the consonant is sometimes doubled before adding the suffix.
>
> if the suffix begins with "i":
>
> > If the stem ends in "ie", this is replaced by "y"
> >
> > If the stem ends in "ue" or "e" preceded by a consonant, then the "e" is dropped
>
> otherwise if the stem ends in "y" preceded by a consonant then the "y" is replaced by "i"

otherwise if the stem ends with "e" and either the suffix starts with "e" or the "e" at the end of the stem is preceded by a consonant or a "u", then the "e" is dropped

To add a suffix beginning with a consonant to a stem:
if the stem ends in "e", then the e *may* be dropped before adding the suffix.

if the stem ends in "y" preceded by a consonant, and the stem is not monosyllabic, then the "y" must be changed to an "i" before adding the suffix.

General suffix stripping rules

NB For these rules "y" is treated as a vowel

To remove a suffix beginning with a vowel:

if the stem after removing the suffix ends in a double consonant, excluding "w" and "x", preceded by a single vowel (or vowel preceded by "qu"), unless the stem ends in "err", "orr" or "omm", one of the consonants is sometimes removed.

if the suffix begins with "i":
If the stem, after removing the suffix ends in "y", this may be replaced by "ie"

If the stem, after removing the suffix ends in "u" or a consonant, then an "e" may be added to the stem

otherwise if the stem ends in "i" preceded by a consonant then the "i" is replaced by "y"

otherwise if either the suffix starts with "e" or the "e" at the end of the stem ends with a consonant or a "u", then an "e" may be added to the stem

To remove a suffix beginning with a consonant to a stem:
an "e" may be added to the stem.

if the stem ends in "i" preceded by a consonant, and the stem is not monosyllabic, then the "i" must be changed to an "y" before adding the suffix.

Abbreviation rules

A word may be formed by abbreviation or another word.

Rules for POS transfer without modification

[A noun may be used as a verb]

[A verb may be used as a noun.]

A verb ending in "-ate" may also exist as an adjective and/or noun.

An adjective of verbal origin ending in"-nt" may also be used as a verb.

Participle rules

The active participle of a verb may be used as an adjective, implying that the noun or pronoun which the adjectival participle qualifies is the subject of the verb whose participle is used adjectivally at the time indicated by the tense of the verb of which the noun or pronoun is an argument.

The passive participle of a verb may be used as an adjective, implying that the noun or pronoun which the adjectival participle qualifies is or was the object of the verb whose participle is used adjectivally at **or before** the time indicated by the tense of the verb of which the noun or pronoun is an argument.

A gerund, morphologically identical to the active participle of a verb, may be used as a noun meaning the process, state or event to which the verb refers.

A passive participle used as an adjective may also be used as a noun, meaning the set of beings or objects to which the adjectival participle could be applied..

If there is an irregular verb in"-t" then there may be an obsolete passive participle with the same form in"-t" still used as an adjective with the same meaning as the adjectival use of the current passive participle of the irregular verb.

Adjective to adverb transformation rules

In all cases the transformation implies that the adjective is applicable to the logical subject of the verb qualified by the adverb, where logical subject means the grammatical subject in the case of an active verb, or a noun governed by the preposition "by" (if any) in the case of a passive verb.

An adverb can be formed from an adjective by adding"-ly".

An adjective may be usable as an adverb without any suffix.

If there is an adjective in"-ic", then the adverb formed from it will be in "-ically" even if there is no form "-ical".

If there is an adjective in "-ble", then the adverb formed from it will be in "-bly".

Verb to adjective transformation rules

If a verb is derived from French, then there may be an adjective formed by appending the suffix "-ant". The meaning of the adjective corresponds to the adjectival use of the active participle.

If a verb is derived directly from Latin, then there may be an adjective of the same form as the stem of the genitive of the Latin present active participle. The meaning of the adjective corresponds to the adjectival use of the active participle.

An adjective in "-ant" derived from a French verb may be imported where no corresponding verb exists in English. The meaning may or may not be the same in the two languages.

[*There may be an adjective formed by adding "-e" to the stem of a Latin passive participle.* If an English verb ending in"-e" has been derived *through French from that Latin passive participle,* then the same adjective may be formed by replacing the"-e" with "-ite". The meaning will be that of the adjectival use of the passive participle of *either* the *Latin, French or* English verb.]

If a verb ends in "-ate", there may be a corresponding adjective ending in "-ative", whose meaning corresponds to the adjectival use of the active participle.

If there is a verb of Latin origin, there may be an adjective in "-ive" formed from the Latin passive participle. The meaning will be that of the adjectival use of the passive participle of either the Latin or the English verb.

An adjective in "-ive" may be formed from the passive participle of a Latin verb even when there is no corresponding verb in English. The meaning is likely to be that of the adjectival use of the passive participle of the Latin verb.

[An adjective may be formed by adding "-ive" to the English verb stem. The meaning is likely to be that of the adjectival use of the active participle of the verb.]

Given a verb in "-ate" *derived from the Latin passive participle in "-atus",* there may also be an adjective in "-ate" *which retains the meaning of the Latin participle.*

[If there is a verb **v** in "-ate" there may be a corresponding adjective in "-able", meaning able to be **v**-ed.]

If there is a verb **v** not ending in "-ate" there may be a corresponding adjective formed by appending "-able", meaning able to be **v**-ed.

If there is a verb in "-ate" there may be a corresponding adjective in "-ative", corresponding to the adjectival use of the active participle of the verb.

[*If a verb **v** is of Latin origin,* there may be an adjective formed by appending "-ible" *to either the Latin infinitive stem or the Latin passive participle stem,* or to the English verb. The meaning is likely to be able to be **v**-ed]

*An adjective in "-ible" may be formed from the passive participle of a Latin verb **v** even when there is no corresponding verb in English. The meaning is likely to be able to be **v**-ed.*

Even if a verb is not derived from Latin, there may be a corresponding adjective by appending "-atious". The meaning is likely to be that of the adjectival use of the active participle with an implication of continuity or repetition.

There may be an adjective formed by appending "-some" to a verb. The meaning is likely to be that of the adjectival use of the active or passive participle with an implication of continuity or repetition.

There may be an adjective formed by appending "-ful" to a verb. The meaning is likely to be that of the adjectival use of the active or passive participle with an implication of continuity or repetition.
There may also be an adjective with a negative meaning formed by appending "-less" to the verb. If both exist, then they are likely to be opposites.

If there is a verb in "-ise"/"-ize" there may be a corresponding adjective in"-ic". (Insufficient examples to determine meaning).

[An adjective may be formed by appending "-ous" to a verb. The meaning is likely to be that of the adjectival use of the active participle with an implication of continuity or repetition.]

An adjective may be formed by appending "-ative" to a verb even where there is no corresponding verb form in "-ate". The meaning is likely to be that of the adjectival use of the active participle.

Verb to noun transformation rules

A noun may be formed from a verb in "-ate" by appending the suffix"-or". The meaning of this noun can correspond to any thematic role performed by the grammatical subject of the verb.

If a verb is formed from a Latin passive participle, then a noun may be formed by appending"-or" to the stem of the Latin passive participle. The meaning of this noun can correspond to any thematic role performed by the grammatical subject of the verb. If the English verb ends in t then the noun may be derived by appending"-or".

[A noun may be formed from a verb *of French origin* by appending the suffix"-or". The meaning of this noun can correspond to any thematic role performed by the grammatical subject of the verb.]

[A noun may be formed from a verb by appending the suffix"-er". The meaning of this noun can correspond to any thematic role performed by the grammatical subject of the verb.]

[If there is a noun formed by appending"-er" to a verb to correspond to its grammatical subject, there may be another noun formed by appending y to the"-er", indicating the result of the verb performed by the noun in"-er" as its grammatical subject.]

A noun may be formed from a verb by appending the suffix"-ee". The meaning of this noun can correspond to any thematic role performed by the grammatical object (direct or indirect) of the verb.

If there is a verb in"-er", there may be a corresponding noun in"-ry", whose meaning is that of the gerund.

[Even if there is no adjective in"-nt" formed from the above rules then there still may be a noun in "-nce" formed as if the adjective in"-nt" existed, whose meaning is that of the gerund.]

If there is a verb in"-er", there may be a corresponding noun in "-rance", whose meaning is that of the gerund.

If there is a verb in"-fy" there may be a corresponding noun in "-fication", whose meaning is that of the gerund.

Given a verb in "-ate" *derived from the Latin passive participle in "-atus",* there may also be a noun in "-ate" *which has the meaning of the result of the Latin verb.*

If a verb **v** ends in"-te", then there may be a corresponding noun ending in "-tion", whose meaning may correspond to the process of **v**-ing, or to the subject of **v**.

If there is a verb of direct or indirect Latin origin, there may be a corresponding noun formed by adding "-ion" to the stem of the Latin passive participle, whose meaning is that of the gerund of either the Latin or the English verb.

If an English verb is formed from the stem of the Latin passive participle, then a noun may formed by adding "-ion" to the English verb if it ends in t *or by adding "-ion" to the stem of the Latin passive participle, whose meaning is that of the gerund of either the Latin or the English verb*

A noun in "-ion" may be formed from the passive participle of a Latin verb even when there is no corresponding verb in English, whose meaning is that of the gerund of either the Latin verb

Even if a verb is not derived from Latin, there may be a corresponding noun formed by appending "-ation", whose meaning is that of the gerund.

If there is a verb in "-ise" there may be a corresponding noun in "-isation" (or "-ize"; "-ization") , whose meaning is that of the gerund.

If there is a verb *derived from Latin through French,* which ends in "-ise", there may be a corresponding noun in "-ice", whose meaning corresponds to the object of the verb.

A noun in "-ism" may be formed from a verb **v** in "-ise" meaning belief in the virtue of **v**-ing.

A noun may be formed by appending "-ist" to a verb **v**, meaning a practitioner or believer in the virtue of **v**-ing.

If there is a verb *of French origin,* there may be a noun in "-age" formed from it, whose meaning is that of the gerund

[There may be a noun formed by adding "-al" to the stem of a verb. Its meaning is likely to correspond to the gerund or to the result of the verb.]

A noun may be formed by adding the suffix "-ment" to a verb. The meaning of the noun may correspond to the meaning of the gerund or the result of the verb.

If there is a verb in "-er" there may be a corresponding noun in "-ery", whose meaning is that of the gerund.

If there is a verb *of French origin* in "-ain", there may be a corresponding noun in "-aint", whose meaning is that of the gerund.

If there is a verb *of Greek origin* in "-yse" then there may be a corresponding noun in "-ysis", whose meaning is that of the gerund

If there is a verb *of Greek origin in "-yse"* then there may be a corresponding noun in "-ysate", whose meaning is that of the object or result of the verb.

Adjective to noun transformation rules

If there is an adjective **j**, ending in"-nt", then there may be a corresponding noun ending in "-nce", whose meaning corresponds to the state of being **j**.

If there is an adjective **j**, ending in"-nt", then there may be a corresponding noun ending in "-ncy", whose meaning corresponds to the state of being **j**.

An adjective formed from a Latin, French or English active participle may also be used as a noun meaning a person with the quality expressed by the adjective.

If there is an adjective ending in "-able" then there may be a corresponding noun ending in "-ability", whose meaning corresponds to the state of being.

If there is an adjective in "-ible", there may be a corresponding noun in "-ibility", whose meaning corresponds to the state of being

If there is an adjective in "-ile" there may be a corresponding noun in "-itility", whose meaning corresponds to the state of being

If there is a adjective in "-ous", there may be a corresponding noun in "-ity", whose meaning corresponds to the state of being.

If there is an adjective in"-al", there may be a corresponding noun in "-ality", whose meaning corresponds to the state of being.

If there is an adjective *of French origin*, there may be a noun formed from it by appending "-ity", whose meaning corresponds to the state of being.

If there is an adjective **j** in "-graphic". There may be a corresponding noun in "-grapher" meaning a person who engages in the study of that which is **j**.

Given an adjective **j**, there may be a noun formed by adding "-ness", meaning the state of being **j**, especially if the adjective ends in "-ous" or "-able".

There may be a noun formed by appending "-ism" to a corresponding adjective **j**, meaning belief in the virtue of being **j** or the state of being **j**.

A noun may be formed by appending "-ist" to an adjective **j** , meaning someone who is or believes in the virtue of beng **j**.

An adjective **j** in "-ive" may also be used as a noun meaning something which is **j**.

If there is an adjective **j** ending in"-te" there may be a corresponding noun in "-tion" meaning something which is **j**.

If there is an adjective **j** in "-ic", there may be a noun in "-ics" formed from it, meaning either the set of things which are **j** or the study of things which are **j**.

An adjective **j** in "-ical" may also be used as a noun, meaning something which is **j**.

An adjective in "-atory" may also be used as a noun with a different meaning.

If there is an adjective in"-e", there may be a corresponding noun in "-ety", whose meaning corresponds to the state of being.

An adjective of Italian origin indicating the manner in which a piece of music is to be played may also be used as a noun referring to the same piece of music.

Noun to adjective transformation rules

An adjective may be formed from a noun by adding "-y". If the noun ends in"-e" then the "-e" may be dropped. The adjective may mean having 1 or more of the noun.

If there is a noun in "-nce" there may be a corresponding adjective in "-ntial", meaning pertaining to or having the characteristic property of the noun.

If there is a noun in"-nt" there may be a corresponding adjective in "-ntial", meaning pertaining to or having the characteristic property of the noun

If there is a noun **n** ending in "-ion", then there may be an adjective ending in "-ional" meaning pertaining to **n**.

An adjective may be formed from a noun in "-ion" by replacing "-ion" with "-ory", meaning pertaining to or having the characteristic property of the noun.

An adjective may be formed from a noun in "-ion" by replacing "-ion" with "-ive", meaning pertaining to or having the characteristic property of the noun.

An adjective may be formed by adding "-ary" to a noun, especially if the noun ends in "-ent" or "-ion", meaning pertaining to or having the characteristic property of the noun

[There may be an adjective formed by adding "-al" to a noun, especially if the noun ends in "-ion", "-our", "-oid", meaning pertaining to or having the characteristic property of the noun.]

If there is a noun **n** ending in "-ist", then there may be an adjective ending in "-istic" meaning the quality of being an **n**.

If there is a noun ending in"-ic" or "-ics", there may be a corresponding adjective in "-ical", meaning pertaining to or having the characteristic property of the noun.

An adjective may be formed by appending "-oid" to a noun **n**, meaning resembling **n** while not being **n**.

If a noun ends in"-y" there may be a corresponding adjective in"-ic" and/or "-ical", meaning pertaining to or having the characteristic property of the noun

[If a noun ends in"-y" there may be a corresponding adjective in"-al" , meaning pertaining to or having the characteristic property of the noun.]

There may be an adjective formed by adding"-ic" to a noun, meaning pertaining to or having the characteristic property of the noun.

[An adjective may be formed by appending "-ous" to a noun. If the noun ends in l, then the l may optionally be doubled, meaning pertaining to or having the characteristic property of the noun.]

[If there is a noun ending in"-y", there may be a corresponding adjective in "-ous" or "-ious", meaning pertaining to or having the characteristic property of the noun.]

If there is an noun *of French origin* ending in"-e", there may be a adjective formed from it by replacing"-e" with "-ious", meaning pertaining to either the French or the English noun.

There may be an adjective formed by appending "-ful" to a noun **n**, meaning full of **n**. There may also be an adjective with a negative meaning formed by appending "-less" to the noun. If both exist, then they are likely to be opposites.

An adjective may be formed by appending "-ic" or "-al" to the genitive stem of a Latin noun. If both exist, they are likely to represent distinct but related meanings.

If there is a noun in"-le" *derived from a Latin noun in "-ulus", "-ula" or "-ulum"* then there may be an adjective in "-ular", meaning pertaining to or having the characteristic property of the noun.

If there is a noun *of Greek origin* ending in "-m" or "-ma", there may be a corresponding adjective in "-matic", meaning pertaining to or having the characteristic property of the noun

If there is a noun in "-nce" there may be a corresponding adjective in "-ncial", meaning pertaining to or having the characteristic property of the noun.

An adjective may be formed by appending "-ed" to a noun **n**, meaning having 1 or more **n**(s).

A noun **n** in "-ist" may also be used as an adjective meaning that the noun qualified by the adjective is also an **n**.

An adjective may be formed from a noun in"-e" by appending"-ly". The adjective may mean having 1 or more of the noun or having the characteristic property of the noun.

There may be an adjective formed by appending "-some" to a noun The adjective is likely to mean having the characteristic property of the noun..

If there is a Latin or Greek word used in the unmodified original nominative for a bodypart, there may be a corresponding adjective formed by appending "-eal" to the genitive stem of the Greek or Latin word , meaning pertaining to or having the characteristic property of the noun.

Noun to verb transformation rules

If a noun **n** ends in"-y" there may be a corresponding verb in "-ise"/"-ize", meaning to practice **n**.

A verb may be formed by appending "-ise" to a noun **n**, meaning cause to become **n**.

A verb may be formed by appending "-en" to a corresponding noun, meaning to add n to the object of the verb.

[There may be a verb formed by appending "-ate" to a noun **n**, meaning to apply **n**.]

If there is a noun **n** in "-nce" there may be a corresponding verb in "-ntiate", meaning to make or show **n**.

If there is a noun **n** in "-e", there may be a related verb in "-ify", meaning to be, become or cause to become **n**.

Adjective to verb transformation rules

A verb may be formed by appending "-ise" to an adjective **j** ending in "-al", meaning cause to become **j**.

A verb may be formed by appending "-ise" to a adjective **j**, meaning cause to become **j**.

If there is a adjective **j** in "-nt" there may be a corresponding verb in "-ntiate", meaning to cause the object of the verb to become or to show the object of the verb to be **j**.

There may be a verb formed by appending "-en" to an adjective **j**, meaning to become or cause to become **j**.

Adverb to adverb transformation rules

An adverb in "-ward" may also be spelt "-wards", without change in meaning.

Adjective to adjective transformation rules

If there is an adjective ending in "-ic", there may be another adjective in "-ical", with the same meaning.

An adjective may exist identical in form to an adverb in "-ly" even though the adjective from which the adverb is derived also exists. There may be a subtle difference in meaning between the two adjectives.

If there is a Latin adjective in "-ilis" there may be a corresponding English adjective in "-ile" with similar meaning.

If there is and an adjective in "-ant" derived from a verb in "-ate" and also another adjective formed by applying a prefix to the first adjective, then there may also be a corresponding adjective with the same prefix but with suffix "-able". The meaning is not established.

If there is an adjective ending in "-te" there may be another adjective in "-tive" with different meaning.

There may be an adjective formed by appending "-ant" to another adjective, having a slightly different meaning.

If there is an adjective *of French origin* ending in "-e", then there may be another adjective with similar meaning ending in "-eous".

If there is an adjective in "-ate", there may be another adjective in "-al" with similar meaning.

Verb to verb transformation rules

If there is an adjective in "-ant" derived from the active participle of a French verb there may be corresponding verb in "-ate" formed from the passive participle of the Latin verb from which the French verb is derived. The second verb is likely to indicate a repetition of the first

If a verb has been derived from Latin through French there may be another verb in "-ate" formed from the Latin passive participle in "-atus". The 2 verbs may have

different shades of meaning. If the first verb ends in "-e", then the second verb may be formed by replacing "-e" with "-ate"

If a verb is derived from the Latin passive participle not ending in "-atus", there may be another verb derived from the Latin passive participle of the iterative form in "-atus". The 2 verbs may have different shades of meaning.

A verb in "-ise" may also be spelt "-ize" with identical meaning.

If there is a verb *of Greek origin* in "-yse" then it may also be spelt "-yze" with identical meaning.

Given a verb ending in "-l" then another l may be added with identical meaning.

Noun to noun transformation rules

If there is a noun **n** ending in "-ic" or "-ics", there may be a corresponding noun in "-icist" meaning a practitioner of **n**.

If there is a Latin or Greek word used in the unmodified original nominative for a bodypart, there may be a corresponding noun formed by appending "-itis" to the genitive stem of the Greek or Latin word, meaning a disease afflicting that bodypart.

There may be a noun **n** formed by adding "-ism" to another noun, meaning the study of or belief in **n**.

A noun in "-i" may also be spelt with "-y" with identical meaning.

If there is a noun **n** in "-ism", there may be another noun in "-ist" meaning a believer in or practitioner of **n**, or vice versa.

There may be a noun formed by appending "-ship" to another noun **n**. The noun in "-ship" is likely to mean the state or status of being an **n**.

An English noun may be formed by removing "-is" from a Latin noun. The English noun may or may not have the same meaning as the Latin noun.

A noun may be formed by appending "-ist" to another noun **n**, meaning a believer in the value of **n**.

[If there is a noun in "-ine", "-ine" may be abbreviated to "-in" with identical meaning.]

If there is a noun in "-nce" there may be a corresponding noun in "-ntial" with a different but related meaning.

[A noun may be formed by appending "-ry" to another noun. There will be a significant difference in meaning.]

[A noun may be formed by appending "-age" to another noun. The meaning will be more abstract.]

There may be an noun formed by appending "-ful" to another noun **n**. Its meaning will be an amount of something contained or borne by **n**.

A noun may be formed by appending "-oid" to another noun **n**, meaning something which resembles n while not being **n**.

A noun in "-y" may also be spelt "-ie" with identical meaning

A noun may be formed by appending "-eer" to another noun **n**. The meaning will be a practitioner of or expert in making or interacting with **n**(s).

If there is a noun **n** ending in "-ty", there may be another noun in "-tarian" meaning a believer in or practitioner of **n**.

A noun may be formed by adding "-ary" to another noun ending in "-ion", meaning a believer in or practitioner of **n**.

[A noun may be formed by appending "-man" to another noun **n** meaning a man who is concerned with **n**.]

Appendix 10

Original table of morphological rules (original version; §3)

Italics in the following table indicate a multilingual rule which was not been implemented. *All morphemes referred to are suffixes*.

Rule				Relation
Source		Target		
Morpheme to remove	POS	Morpheme to append	POS	
	VERB	ing	ADJECTIVE	Participle
	VERB	ed	ADJECTIVE	Participle
	VERB	ing	NOUN	Gerund
	VERB	ed	NOUN	Gerund
t	VERB	t	ADJECTIVE	Participle
	ADJECTIVE	ly	ADVERB	Pertainym
	ADJECTIVE		ADVERB	Pertainym
ic	ADJECTIVE	ically	ADVERB	Pertainym
ble	ADJECTIVE	bly	ADVERB	Pertainym
	VERB	ant	ADJECTIVE	Participle
ans	*LATIN ACTIVE PARTICIPLE*	*ant*	*ADJECTIVE*	*Participle*
ens	*LATIN ACTIVE PARTICIPLE*	*ent*	*ADJECTIVE*	
ant	*FRENCH ACTIVE PARTICIPLE*	ant	ADJECTIVE	Participle
us	*LATIN PASSIVE PARTICIPLE*	*e*	*ADJECTIVE*	*Participle*
e	VERB	ite	ADJECTIVE	
ate	VERB	ative	ADJECTIVE	Participle
us	*LATIN PASSIVE PARTICIPLE*	*ive*	*ADJECTIVE*	*Participle*
us	*LATIN PASSIVE PARTICIPLE*	*ive*	*ADJECTIVE*	Participle
	VERB	ive	ADJECTIVE	Participle
ate	VERB	ate	ADJECTIVE	Participle
ate	VERB	able	ADJECTIVE	Potential
	VERB	able	ADJECTIVE	Potential
ate	VERB	ative	ADJECTIVE	Participle
are	*LATIN INFINITIVE*	*ible*	*ADJECTIVE*	*Potential*
ere	*LATIN INFINITIVE*	*ible*	*ADJECTIVE*	
ire	*LATIN INFINITIVE*	*ible*	*ADJECTIVE*	
us	*LATIN PASSIVE PARTICIPLE*	*ible*	*ADJECTIVE*	
	VERB	ible	ADJECTIVE	
us	*LATIN PASSIVE PARTICIPLE*	*ible*	*ADJECTIVE*	*Potential*
	VERB	atious	ADJECTIVE	Participle
	VERB	some	ADJECTIVE	Participle

Rule				Relation
Source		Target		
Morpheme to remove	**POS**	**Morpheme to append**	**POS**	
	VERB	ful	ADJECTIVE	Participle
	VERB	less	ADJECTIVE	Antonym of above
ise	VERB	ic	ADJECTIVE	Indeterminate
ize	VERB	ic	ADJECTIVE	
	VERB	ous	ADJECTIVE	Participle
	VERB	ative	ADJECTIVE	Participle
	VERB		NOUN	Indeterminate
ate	VERB	ator	NOUN	Subject
tus	*LATIN PASSIVE PARTICIPLE*	*tor*	*NOUN*	*Subject*
t	VERB	tor	NOUN	
	VERB	or	NOUN	Subject
	VERB	er	NOUN	Subject
	VERB	ee	NOUN	Object
er	VERB	ry	NOUN	Gerund
nt	VERB	nce	NOUN	Gerund
	VERB	ance	NOUN	
er	VERB	rance	NOUN	Gerund
fy	VERB	fication	NOUN	Gerund
ate	VERB	ate	NOUN	Result
te	VERB	tion	NOUN	Gerund
				Subject
us	*LATIN PASSIVE PARTICIPLE*	*ion*	*NOUN*	*Gerund*
te	VERB	tion	NOUN	Gerund
us	*LATIN PASSIVE PARTICIPLE*	*ion*	*NOUN*	*Gerund*
ise	VERB	ation	NOUN	Gerund
ise	VERB	isation	NOUN	Gerund
ize	VERB	ization	NOUN	
ise	VERB	ice	NOUN	Object
ise	VERB	ism	NOUN	Belief/practice
	VERB	ist	NOUN	Believer/practioner
	VERB	age	NOUN	Gerund
	VERB	al	NOUN	Gerund
				Result
er	VERB	ment	NOUN	Gerund
				Result
er	VERB	ery	NOUN	Gerund
				Result
ain	VERB	aint	NOUN	Gerund
yse	VERB	ysis	NOUN	Gerund
yse	VERB	ysate	NOUN	Object
				Result
nt	ADJECTIVE	nce	NOUN	StateOfBeing
nt	ADJECTIVE	ncy	NOUN	StateOfBeing

Rule				Relation
Source		Target		
Morpheme to remove	POS	Morpheme to append	POS	
nt	ADJECTIVE	nt	NOUN	Qualified
able	ADJECTIVE	ability	NOUN	StateOfBeing
ible	ADJECTIVE	ibility	NOUN	StateOfBeing
ile	ADJECTIVE	itility	NOUN	StateOfBeing
ous	ADJECTIVE	ity	NOUN	StateOfBeing
al	ADJECTIVE	ality	NOUN	StateOfBeing
	ADJECTIVE	ity	NOUN	StateOfBeing
graphic	ADJECTIVE	grapher	NOUN	ScholarOfThatWhichIs
	ADJECTIVE	ness	NOUN	StateOfBeing
	ADJECTIVE	ism	NOUN	Belief/practice
	ADJECTIVE	ist	NOUN	Believer/practioner
ive	ADJECTIVE	ive	NOUN	Qualified
te	ADJECTIVE	tion	NOUN	Qualified
ic	ADJECTIVE	ics	NOUN	Qualified
				ScholarOfThatWhichIs
ical	ADJECTIVE	ical	NOUN	Qualified
atory	ADJECTIVE	atory	NOUN	Indeterminate
e	ADJECTIVE	ety	NOUN	StateOfBeing
	ADJECTIVE		NOUN	Qualified
	NOUN	y	ADJECTIVE	Having
e	NOUN	y	ADJECTIVE	
nce	NOUN	ntial	ADJECTIVE	Pertainym
				ChacterisedBy
nt	NOUN	ntial	ADJECTIVE	Pertainym
				ChacterisedBy
ion	NOUN	ional	ADJECTIVE	Pertainym
ion	NOUN	ory	ADJECTIVE	Pertainym
				ChacterisedBy
ion	NOUN	ive	ADJECTIVE	Pertainym
				ChacterisedBy
ent	NOUN	entary	ADJECTIVE	Pertainym
ion	NOUN	ionary	ADJECTIVE	ChacterisedBy
	NOUN	al	ADJECTIVE	Pertainym
				ChacterisedBy
ist	NOUN	istic	ADJECTIVE	BeingA
ic	NOUN	ical	ADJECTIVE	Pertainym
ics	NOUN	ical	ADJECTIVE	ChacterisedBy
	NOUN	oid	ADJECTIVE	Resembling
y	NOUN	ic	ADJECTIVE	Pertainym
y	NOUN	al	ADJECTIVE	ChacterisedBy
y	NOUN	ical	ADJECTIVE	
	NOUN	ic	ADJECTIVE	Pertainym
				ChacterisedBy

Rule				Relation
Source		Target		
Morpheme to remove	**POS**	**Morpheme to append**	**POS**	
	NOUN	ous	ADJECTIVE	Pertainym
				ChacterisedBy
y	NOUN	ous	ADJECTIVE	Pertainym
y	NOUN	ious	ADJECTIVE	ChacterisedBy
e	NOUN	ious	ADJECTIVE	Pertainym
	NOUN	ful	ADJECTIVE	Having
	NOUN	less	ADJECTIVE	Antonym of above
is	*LATIN GENITIVE*	*ic*	*ADJECTIVE*	*Indeterminate*
is	*LATIN GENITIVE*	*al*	*ADJECTIVE*	
le	NOUN	ular	ADJECTIVE	Pertainym
				ChacterisedBy
m	NOUN	matic	ADJECTIVE	Pertainym
ma	NOUN	matic	ADJECTIVE	ChacterisedBy
nce	NOUN	ncial	ADJECTIVE	Pertainym
				ChacterisedBy
	NOUN	ed	ADJECTIVE	Having
ist	NOUN	ist	ADJECTIVE	BeingA
e	NOUN	ely	ADJECTIVE	Having
				ChacterisedBy
	NOUN	some	ADJECTIVE	ChacterisedBy
is	LATIN GENITIVE	eal	ADJECTIVE	Pertainym
os	GREEK GENITIVE	eal	ADJECTIVE	Pertainym
	NOUN		VERB	Indeterminate
y	NOUN	ise	VERB	Practice
y	NOUN	ize	VERB	
	NOUN	ise	VERB	Make
	NOUN	ize	VERB	
	NOUN	en	VERB	AddTo
	NOUN	ate	VERB	Make
				AddTo
nce	NOUN	ntiate	VERB	Show
e	NOUN	ify	VERB	Make
				Become
al	ADJECTIVE	alise	VERB	Make
al	ADJECTIVE	alize	VERB	
	ADJECTIVE	ise	VERB	Make
nt	ADJECTIVE	ntiate	VERB	Make
				Show
	ADJECTIVE	en	VERB	Make
				Become
ward	ADVERB	wards	ADVERB	Synonym
ic	ADJECTIVE	ical	ADJECTIVE	Synonym

Rule				Relation
Source		Target		
Morpheme to remove	**POS**	**Morpheme to append**	**POS**	
	ADJECTIVE	ly	ADJECTIVE	NearSynonym
ilis	*LATIN ADJECTIVE*	*ile*	*ADJECTIVE*	*NearSynonym*
ant	ADJECTIVE	able	ADJECTIVE	Indeterminate
te	ADJECTIVE	tive	ADJECTIVE	Indeterminate
	ADJECTIVE	ant	ADJECTIVE	NearSynonym
e	ADJECTIVE	eous	ADJECTIVE	NearSynonym
ate	ADJECTIVE	al	ADJECTIVE	NearSynonym
al	ADJECTIVE	ate	ADJECTIVE	
atus	*LATIN PASSIVE PARTICIPLE*	*ate*	*VERB*	*IterationOf*
e	VERB	ate	VERB	NearSynonym
us	LATIN PASSIVE PARTICIPLE	ate	VERB	NearSynonym
ise	VERB	ize	VERB	Synonym
yse	VERB	yze	VERB	Synonym
l	VERB	ll	VERB	Synonym
ics	NOUN	icist	NOUN	Believer/practioner
is	LATIN GENITIVE	itis	NOUN	AfflictionOf
os	GREEK GENITIVE	itis	NOUN	AfflictionOf
	NOUN	ism	NOUN	Belief/practice
i	NOUN	y	NOUN	Synonym
ism	NOUN	ist	NOUN	Believer/practioner
ist	NOUN	ism	NOUN	Belief/practice
	NOUN	ship	NOUN	StateOfBeing
is	*LATIN NOUN*		*NOUN*	*Indeterminate*
	NOUN	ist	NOUN	Believer/practioner
ine	NOUN	in	NOUN	Synonym
nce	NOUN	ntial	NOUN	Indeterminate
	NOUN	ry	NOUN	Indeterminate
	NOUN	age	NOUN	Indeterminate
	NOUN	ful	NOUN	MeasuredBy
	NOUN	oid	NOUN	Resembling
y	NOUN	ie	NOUN	Synonym
	NOUN	eer	NOUN	Believer/practioner
ty	NOUN	tarian	NOUN	Believer/practioner
ion	NOUN	ionary	NOUN	Believer/practioner
	NOUN	man	NOUN	Pertainym
				Believer/practioner
				PurveyorOf
				Indeterminate

Appendix 11

Words autogenerated from CatVar headwords but unrelated to them

chancery	NOUN
cursive	NOUN
cursive	ADJECTIVE
cursively	ADVERB
cursor	NOUN
cursorily	ADVERB
cursory	ADJECTIVE
fallal	NOUN
fallibility	NOUN
fallible	ADJECTIVE
fellate	VERB
fellation	NOUN
feller	NOUN
fin	NOUN
fin	VERB
final	NOUN
final	ADJECTIVE
finalisation	NOUN
finalise	VERB
finalist	NOUN
finality	NOUN
finalization	NOUN
finalize	VERB
finally	ADVERB
finance	NOUN
finance	VERB
financial	ADJECTIVE
financially	ADVERB
financing	NOUN
finite	ADJECTIVE
finitely	ADVERB
finiteness	NOUN
finned	NOUN
finned	ADJECTIVE
finning	NOUN
finning	ADJECTIVE
forage	NOUN
forage	VERB
forager	NOUN
foraging	NOUN
lacerate	VERB
lacerate	ADJECTIVE
lacerated	ADJECTIVE
laceration	NOUN
mater	NOUN
matman	NOUN

moral	ADJECTIVE
moralisation	NOUN
moralise	VERB
moralism	NOUN
moralist	NOUN
moralistic	ADJECTIVE
morality	NOUN
moralization	NOUN
moralize	VERB
moralizing	NOUN
morally	ADVERB
pilous	ADJECTIVE
probability	NOUN
probable	ADJECTIVE
probably	ADVERB
pursy	ADJECTIVE
readily	ADVERB
readiness	NOUN
ready	ADJECTIVE
squash	NOUN
still	NOUN
still	VERB
tier	NOUN
tiered	ADJECTIVE

Appendix 12

Productivity of morphological rules (CatVar dataset)

				Full ruleset	Restricted ruleset	Full ruleset
Source		Target		Lexically valid execs.	Lexically valid execs.	
Word Form	POS	Word Form	POS			Total overgen.
	N		V	220	n/a	4
	V		N	219	n/a	1
	Adj.	ly	Adv.	149	130	0
	V	ed	Adj.	133	129	0
	V	er	N	126	n/a	4
	V	ing	N	113	108	0
	Adj.	ness	N	100	88	0
	N	ed	Adj.	90	89	0
	V	ing	Adj.	64	60	0
te	V	tion	N	45	12	0
	V	ation	N	44	37	0
	Adj.	ity	N	37	34	0
	N	y	Adj.	31	n/a	4
ise	V	ize	V	28	25	0
	V	able	Adj.	27	27	0
	Adj.		Adv.	27	27	0
	N	al	Adj.	26	n/a	8

Source		Target		Full ruleset	Restricted ruleset	Full ruleset
Word Form	POS	Word Form	POS	Lexically valid execs.	Lexically valid execs.	Total overgen.
	V	ive	Adj.	26	n/a	3
	V	or	N	26	n/a	3
ion	N	ive	Adj.	25	1	0
	N	ic	Adj.	23	21	0
	V	ment	N	23	23	0
ate	V	ator	N	20	2	0
nt	Adj.	nce	N	19	19	0
ic	Adj.	ical	Adj.	18	1	0
ic	Adj.	ically	Adv.	17	1	0
	N	ise	V	15	14	0
	N	ize	V	15	14	0
	N	ism	N	15	15	0
	N	ist	N	15	13	0
ate	V	ative	Adj.	15	2	0
ate	V	ative	Adj.	15	2	0
te	Adj.	tion	N	15	12	0
	N	less	Adj.	14	14	0
	Adj.	ism	N	14	13	0
	Adj.	ize	V	14	12	0
able	Adj.	ability	N	13	2	0
al	Adj.	ality	N	13	2	0
ble	Adj.	bly	Adv.	12	2	0
nt	Adj.	ncy	N	12	12	0
	N	ous	Adj.	11	n/a	4
	N	man	N	11	n/a	1
ism	N	ist	N	11	9	0
ist	N	istic	Adj.	11	9	0
ist	N	ist	Adj.	11	10	0
	V	less	Adj.	10	10	0
ate	V	ate	Adj.	10	2	0
ate	V	ate	Adj.	10	2	0
ate	V	ate	N	10	2	0
ise	V	isation	N	10	1	0
ize	V	ization	N	10	7	0
nt	Adj.	nt	N	10	12	0
	N	ate	V	9	n/a	6
	V	ist	N	9	9	0
	Adj.	ist	N	9	7	0
ion	N	ional	Adj.	9	1	0
ion	N	ory	Adj.	9	1	0
t	V	tion	N	9	12	0
y	N	ic	Adj.	9	12	0
	V	ent	Adj.	8	8	0
	V	al	N	8	n/a	3
ate	V	able	Adj.	8	n/a	3
e	N	y	Adj.	8	2	0
	N	ful	Adj.	7	7	0
	N	ship	N	7	3	0

				Full ruleset	Restricted ruleset	Full ruleset
Source		Target		Lexically valid execs.	Lexically valid execs.	
Word Form	POS	Word Form	POS			Total overgen.
	V	ful	Adj.	7	7	0
	V	ous	Adj.	7	n/a	1
al	Adj.	alise	V	7	2	0
al	Adj.	alize	V	7	2	0
	N	ry	N	6	n/a	1
	N	age	N	6	n/a	4
al	Adj.	ate	Adj.	6	2	0
	N	en	V	5	5	0
	V	ant	Adj.	5	5	0
ics	N	ical	Adj.	5	1	0
ise	V	ic	Adj.	5	1	0
ise	V	ism	N	5	3	0
ive	Adj.	ive	N	5	3	0
ize	V	ic	Adj.	5	4	0
ize	V	ism	N	5	3	0
y	N	ical	Adj.	5	5	0
	V	ible	Adj.	4	n/a	2
	V	ative	Adj.	4	4	0
	V	ery	N	4	n/a	1
	V	ance	N	4	n/a	4
	V	age	N	4	4	0
	Adj.	en	V	4	4	0
e	V	ate	V	4	2	0
ic	Adj.	ics	N	4	1	0
y	N	ise	V	4	5	0
y	N	ize	V	4	5	0
	V	ee	N	3	3	0
	Adj.	ly	Adj.	3	3	0
fy	V	fication	N	3	1	0
ic	N	ical	Adj.	3	1	0
nce	N	ntial	Adj.	3	3	0
ous	Adj.	ity	N	3	12	0
te	Adj.	tive	Adj.	3	12	0
y	N	ous	Adj.	3	5	0
	V	ed	N	2	2	0
	N	some	Adj.	2	2	0
	N	ful	N	2	2	0
	Adj.	ant	Adj.	2	2	0
ant	Adj.	able	Adj.	2	2	0
e	N	ious	Adj.	2	2	0
e	V	ite	Adj.	2	n/a	3
graphic	Adj.	grapher	N	2	1	0
i	N	y	N	2	1	0
ible	Adj.	ibility	N	2	1	0
ion	N	ionary	Adj.	2	1	0
l	V	ll	V	2	2	0
le	N	ular	Adj.	2	2	0
nt	N	ntial	Adj.	2	2	0

				Full ruleset	Restricted ruleset	Full ruleset
Source		Target		Lexically valid execs.	Lexically valid execs.	
Word Form	POS	Word Form	POS			Total overgen.
ty	N	tarian	N	2	12	0
	N	oid	N	1	1	0
	N	eer	N	1	1	0
	V	atious	Adj.	1	1	0
	V	some	Adj.	1	1	0
ain	V	aint	N	1	2	0
atory	Adj.	atory	N	1	2	0
e	N	ely	Adj.	1	2	0
e	N	ify	V	1	2	0
e	Adj.	ety	N	1	1	0
e	Adj.	eous	Adj.	1	1	0
ent	N	entary	Adj.	1	1	0
er	V	ry	N	1	1	0
er	V	rance	N	1	1	0
er	V	ery	N	1	1	0
ical	Adj.	ical	N	1	1	0
ics	N	icist	N	1	1	0
ine	N	in	N	1	n/a	6
ion	N	ionary	N	1	1	0
ise	V	ice	N	1	1	0
m	N	matic	Adj.	1	1	0
ma	N	matic	Adj.	1	1	0
Ma	N	matise	V	1	1	0
Ma	N	matize	V	1	1	0
Nce	N	ncial	Adj.	1	0	0
Nce	N	ntiate	V	1	1	0
Nce	N	ntial	N	1	1	0
Nt	Adj.	ntiate	V	1	12	0
T	V	tor	N	1	12	0
ward	Adv.	wards	Adv.	1	12	0
Y	N	al	Adj.	1	n/a	11
Y	N	ie	N	1	5	0
Yse	V	ysis	N	1	5	0
Yse	V	ysate	N	1	5	0
yse	V	yze	V	1	5	0
	N	oid	Adj.	0	0	0
ic	N	icist	N	0	1	0
ile	Adj.	itility	N	0	1	0
m	N	matise	V	0	0	0
m	N	matize	V	0	0	0
nt	Adj.	nt	V	0	0	0
				2326	1317	77

Appendix 13

Productivity of morphological rules (Word list dataset)

Source		Target			
Wordform	**POS**	**Wordform**	**POS**	**Lexically valid execs.**	**Total overgeneration**
	VERB		NOUN	176	0
	NOUN		VERB	121	0
	ADJECTIVE	ly	ADVERB	89	1
	ADJECTIVE		ADVERB	66	0
	ADJECTIVE	ness	NOUN	63	1
	VERB	er	NOUN	59	0
	VERB	ing	NOUN	48	0
	VERB	ed	ADJECTIVE	43	1
	NOUN	ed	ADJECTIVE	34	0
	VERB	ing	ADJECTIVE	24	0
	VERB	ation	NOUN	24	0
	NOUN	y	ADJECTIVE	22	3
ise	VERB	ize	VERB	17	0
	NOUN	ic	ADJECTIVE	14	0
ism	NOUN	ist	NOUN	14	0
	VERB	ion	NOUN	13	0
ize	VERB	ization	NOUN	13	0
	NOUN	al	ADJECTIVE	12	0
	NOUN	ist	NOUN	12	0
	NOUN	ism	NOUN	11	0
te	VERB	tion	NOUN	11	0
	ADJECTIVE	ism	NOUN	10	0
	ADJECTIVE	ly	ADJECTIVE	10	0
ic	ADJECTIVE	ical	ADJECTIVE	10	0
ion	NOUN	ive	ADJECTIVE	10	0
ize	VERB	ism	NOUN	9	0
ise	VERB	isation	NOUN	8	0
	NOUN	ship	NOUN	7	1
	NOUN	man	NOUN	7	0
	VERB	al	NOUN	7	0
	VERB	ment	NOUN	7	2
ate	VERB	ate	NOUN	7	0
ise	VERB	ism	NOUN	7	0
	NOUN	ous	ADJECTIVE	6	0
	NOUN	less	ADJECTIVE	6	1
able	ADJECTIVE	ability	NOUN	6	0
ble	ADJECTIVE	bly	ADVERB	6	0
ic	ADJECTIVE	ically	ADVERB	6	0
ion	NOUN	ory	ADJECTIVE	6	0
ive	ADJECTIVE	ive	NOUN	6	0

Source		Target			
Wordform	**POS**	**Wordform**	**POS**	**Lexically valid execs.**	**Total overgeneration**
	NOUN	ise	VERB	5	0
	NOUN	ize	VERB	5	0
	NOUN	ry	NOUN	5	0
	VERB	able	ADJECTIVE	5	0
	VERB	or	NOUN	5	0
	ADJECTIVE	ity	NOUN	5	0
	ADJECTIVE	ize	VERB	5	0
	NOUN	ful	ADJECTIVE	4	0
	VERB	ful	ADJECTIVE	4	0
	VERB	less	ADJECTIVE	4	0
	VERB	ist	NOUN	4	0
	ADJECTIVE	ist	NOUN	4	0
al	ADJECTIVE	alise	VERB	4	0
al	ADJECTIVE	alize	VERB	4	0
ate	VERB	ator	NOUN	4	0
e	NOUN	y	ADJECTIVE	4	3
ion	NOUN	ional	ADJECTIVE	4	0
ise	VERB	ic	ADJECTIVE	4	0
ize	VERB	ic	ADJECTIVE	4	0
nt	ADJECTIVE	nce	NOUN	4	1
nt	ADJECTIVE	ncy	NOUN	4	0
nt	ADJECTIVE	nt	NOUN	4	0
y	NOUN	ic	ADJECTIVE	4	0
	NOUN	ate	VERB	3	0
	NOUN	age	NOUN	3	0
	VERB	ant	ADJECTIVE	3	2
	VERB	ive	ADJECTIVE	3	0
	VERB	ery	NOUN	3	0
	VERB	ance	NOUN	3	0
	VERB	age	NOUN	3	0
	ADJECTIVE	en	VERB	3	0
al	ADJECTIVE	ality	NOUN	3	0
ate	VERB	ative	ADJECTIVE	3	0
ate	VERB	ative	ADJECTIVE	3	0
ist	NOUN	ist	ADJECTIVE	3	0
	NOUN	ful	NOUN	2	0
	NOUN	oid	NOUN	2	0
	VERB	ous	ADJECTIVE	2	0
	VERB	ee	NOUN	2	1
ate	VERB	able	ADJECTIVE	2	0
atory	ADJECTIVE	atory	NOUN	2	0
graphic	ADJECTIVE	grapher	NOUN	2	0
ible	ADJECTIVE	ibility	NOUN	2	0
ist	NOUN	istic	ADJECTIVE	2	0
y	NOUN	ie	NOUN	2	0

Source		Target			
Wordform	**POS**	**Wordform**	**POS**	**Lexically valid execs.**	**Total overgeneration**
	VERB	ed	NOUN	1	1
	NOUN	oid	ADJECTIVE	1	0
	NOUN	en	VERB	1	0
	VERB	some	ADJECTIVE	1	0
	VERB	ative	ADJECTIVE	1	3
	ADJECTIVE	ant	ADJECTIVE	1	0
al	ADJECTIVE	ate	ADJECTIVE	1	1
ate	VERB	ate	ADJECTIVE	1	0
ate	VERB	ate	ADJECTIVE	1	0
e	NOUN	ify	VERB	1	0
er	VERB	ery	NOUN	1	0
ic	NOUN	ical	ADJECTIVE	1	0
ic	NOUN	icist	NOUN	1	0
ical	ADJECTIVE	ical	NOUN	1	0
ics	NOUN	ical	ADJECTIVE	1	0
ine	NOUN	in	NOUN	1	0
ma	NOUN	matic	ADJECTIVE	1	0
nt	ADJECTIVE	nt	VERB	1	0
ous	ADJECTIVE	ity	NOUN	1	0
t	VERB	tion	NOUN	1	0
te	ADJECTIVE	tion	NOUN	1	0
te	ADJECTIVE	tive	ADJECTIVE	1	0
ty	NOUN	tarian	NOUN	1	0
y	NOUN	ical	ADJECTIVE	1	0
y	NOUN	ous	ADJECTIVE	1	0
	NOUN	some	ADJECTIVE	0	0
	NOUN	eer	NOUN	0	0
	VERB	ent	ADJECTIVE	0	0
	VERB	ible	ADJECTIVE	0	0
	VERB	atious	ADJECTIVE	0	0
ain	VERB	aint	NOUN	0	0
ant	ADJECTIVE	able	ADJECTIVE	0	0
e	NOUN	ious	ADJECTIVE	0	0
e	NOUN	ely	ADJECTIVE	0	0
e	VERB	ite	ADJECTIVE	0	0
e	VERB	ate	VERB	0	0
e	ADJECTIVE	ety	NOUN	0	0
e	ADJECTIVE	eous	ADJECTIVE	0	0
ent	NOUN	entary	ADJECTIVE	0	0
er	VERB	ry	NOUN	0	0
er	VERB	rance	NOUN	0	0
fy	VERB	fication	NOUN	0	0
i	NOUN	y	NOUN	0	0
ic	ADJECTIVE	ics	NOUN	0	0
ics	NOUN	icist	NOUN	0	0

Source		Target			
Wordform	**POS**	**Wordform**	**POS**	**Lexically valid execs.**	**Total overgeneration**
ile	ADJECTIVE	itility	NOUN	0	0
ion	NOUN	ionary	ADJECTIVE	0	0
ion	NOUN	ionary	NOUN	0	0
ise	VERB	ice	NOUN	0	0
l	VERB	ll	VERB	0	0
le	NOUN	ular	ADJECTIVE	0	0
m	NOUN	matic	ADJECTIVE	0	0
m	NOUN	matise	VERB	0	0
m	NOUN	matize	VERB	0	0
ma	NOUN	matise	VERB	0	0
ma	NOUN	matize	VERB	0	0
nce	NOUN	ntial	ADJECTIVE	0	0
nce	NOUN	ncial	ADJECTIVE	0	0
nce	NOUN	ntiate	VERB	0	0
nce	NOUN	ntial	NOUN	0	0
nt	NOUN	ntial	ADJECTIVE	0	0
nt	ADJECTIVE	ntiate	VERB	0	0
t	VERB	tor	NOUN	0	0
ward	ADVERB	wards	ADVERB	0	0
y	NOUN	al	ADJECTIVE	0	0
y	NOUN	ise	VERB	0	0
y	NOUN	ize	VERB	0	0
yse	VERB	ysis	NOUN	0	0
yse	VERB	ysate	NOUN	0	0
yse	VERB	yze	VERB	0	0
				1207	22

Appendix 14 Application of generalised spelling rules for suffix stripping

The application of generalised spelling rules by `Suffixer.remove` is applied to a specified original word with a specified original suffix and returns a String array. The algorithm implemented can be represented as follows ('y' is treated as a vowel throughout):

```
if the stem is an empty String then an empty array is returned;
otherwise a default stem is generated by deleting the original suffix
from the end of the original word;
if the original suffix is an empty String then the default stem is
returned, otherwise execution proceeds as follows:
if the original suffix ends with a vowel
{
      if the default stem does not end with 'w', 'x', 'z', 'err',
      'orr' or 'omm' or any vowel, and either the stem ends with a
      double letter or the last 3 letters of the stem are preceded by
```

```
      "qu", then the default stem without its final letter is
      returned followed by the default stem,
      otherwise
      {
            if the default stem ends with 'y' and the original suffix
            stats with 'i' then the default stem is returned followed
            by the stem with "ie" appended
            otherwise,
            {
                  if the default stem ends with 'i' preceded by a
                  consonant then the default stem is returned with
                  'e' appended followed by the default stem
                  otherwise,
                  {
                        if the default stem ends with 'u' or a
                        consonant preceded by any letter
                        {
                              if the default stem ends with 2
                              consonants neither of which is 'w'
                              {
                                    if the default stem ends with 'r,
                                    then the default stem is
                                    returned, followed by the default
                                    stem with 'e' inserted before the
                                    final 'r', followed by the
                                    default stem with 'o' inserted
                                    before the final 'r'
                                    if the default stem ends with
                                    'h', then the default stem is
                                    returned followed by the default
                                    stem with 'e' appended
                                    if the default stem ends with
                                    'c', 's', 'l', 'v' or 'g' NOT
                                    preceded by 'n', then the default
                                    stem is returned with 'e'
                                    appended,
                                    otherwise the default stem is
                                    returned;
                              }
                              otherwise
                              {
                                    if the default stem is
                                    monosyllabic and the last letter
                                    of the default stem is NOT
                                    preceded by 2 vowels and the
                                    default stem does not end with
                                    'x', then the default stem is
                                    returned with 'e' appended
                                    otherwise, the default stem is
                                    returned followed by the default
                                    stem with 'e' appended;
                              }
                        }
                  }
            }
      }
}
if the original suffix ends with a consonant
{
```

```
      if the default stem ends with 'i' and is not monosyllabic and
      the final 'i' is preceded by a consonant, then the default stem
      is returned with the finqal 'i' replaced by 'y,'
      otherwise
      {
            if the original suffix is "s"
            {
                  if the default stem ends with 's', 'z', 'ch' or
                  'zh', then an empty array is returned,
                  otherwise
                  {
                        if the default stem ends with 'e'
                        {
                              if the default stem ends with "se" or
                              "ze", then the default stem with the
                              final 'e' removed is returned, followed
                              by the default stem,
                              otherwise
                              {
                                    if the default stem ends with
                                    "xe", "che" or "zhe", then the
                                    default stem with the final 'e'
                                    removed is returned,
                                    otherwise
                                    {
                                          if the default stem ends
                                          with "ie", then the default
                                          stem is returned with the
                                          final "ie" replaced by 'y',
                                          followed by the default
                                          stem,
                                          otherwise the default stem
                                          is returned;
                                    }
                              }
                        }
                        otherwise the default stem is returned;
                  }
            }
            otherwise
            {
                  if the default stem ends with 'l', then the default
                  stem is returned followed by the default stem with
                  the final 'l' doubled,
                  otherwise the default stem is returned;
            }
      }
}
```

Appendix 15

Undergeneration in suffix stripping (*italics refer to unimplemented multilingual rules*)

Hyper-undergeneration	Undergeneration	Headword	Reason
	lie	lair	Irregular
	cecum	cecal	um->al
	duke	*ducal*	*Asynchronous French imports*

Hyper-undergeneration	Undergeneration	Headword	Reason
	old	older	Adjective comparison (inflectional)
	sand	sands	Plural (inflectional)
	spec	specs	Plural (inflectional)
	ameba	ameban	a->an
	blink	blinks	Plural
	silk	silken	-en
	wool	woolen	-en
	cavalier	*cavalry*	*Asynchronous French imports*
	conceive	*conceit*	*Asynchronous French imports*
draw	drawer	drawers	Plural
	elysium	elysian	um->an
fun	funny	funnies	Plural
	genus	*general*	*Latin genitive*
	inside	insider	POS
	omen	*ominous*	*Latin genitive*
	require	*requite*	*Latin passive participle*
	spark	sparkle	-le
	emerge	*emersion*	*Latin passive participle*
	habit	*habitual*	*Asynchronous French imports*
	judge	*judicial*	*Asynchronous French imports*
	nucleus	nucellus	Irregular
	pretend	*pretence*	*French morphological rule*
	skit	skittish	-ish
ward	warder	wardress	e dropped
	girl	girlish	-ish
	indent	*indenture*	*-ure*
	plenty	plenteous	y->eous
	secede	*secession*	*Latin passive participle*
	serf	*servile*	*French morphological rule*
	solemn	solemness	n dropped
	tomato	*tomatillo*	*Spanish morphological rule*
	velvet	velveteen	-een
	assume	*assumption*	*Latin passive participle*
	deposit	depositary	POS
	forfeit	forfeiture	-ure
	perceive	*perceptual*	*French/Latin derivation*
	pharmacy	pharmacist	y->ist
	vagina	*vaginismus*	*German/Latin derivation*
	approve	*approbate*	*Latin passive*

Hyper-undergeneration	**Undergeneration**	**Headword**	**Reason**
			participle
	bounty	bounteous	y->eous
	exclaim	*exclamation*	*Latin passive participle*
	gas	gaseous	-eous
inherit	inheritor	inheritress	or->ress
	mount	*mountain*	*French morphological rule*
	substance	substantive	nce->ntive
	contempt	contemptuous	-uous
	destroy	*destruct*	*Latin passive participle*
	evolve	*evolution*	*Latin passive participle*
	genus	*generate*	*Latin genitive*
	microphone	microphoning	POS
	orchestra	orchestrate	a->ate
	paradise	paradisaic	Irregular spelling
	prank	prankish	-ish
	register	registration	e dropped
	spermatazoon	spermatozoan	Irregular spelling
	transmit	*transmission*	*Latin passive participle*
	admit	*admissibility*	*Latin passive participle*
	contract	contractual	-ual
	destroy	*destruct*	*Latin passive participle*
	reciprocal	reciprocate	POS
	romance	romantic	ce->tic
	series	*serial*	*Latin morphological rule*
tranquil	tranquilise	tranquilising	not in lexicon
	antithesis	*antithetic*	*Greek genitive*
elect	election	electioneer	POS
	enterprise	enterprising	POS
	permit	*permission*	*Latin passive participle*

Appendix 16

Candidate prefixes

First 100 sorted on heuristic $\frac{f_c^2}{f_p}$

Prefix	f_c	$\frac{f_c}{f_p}$	$\frac{f_c^2}{f_p}$	Semantic validity
un	2227	0.869582	1936.559	Valid
in	1698	0.638826	1084.727	Valid
co	2332	0.37753	880.3989	Valid
re	1543	0.541974	836.2659	Valid
s	6905	0.087115	601.5294	Invalid
de	1340	0.363144	486.6125	Valid
c	6177	0.07793	481.3763	Invalid
di	1212	0.328455	398.0878	Valid
dis	662	0.546205	361.5875	Valid
p	5345	0.067434	360.4333	Invalid
a	4778	0.06028	288.0194	Valid
pro	589	0.487583	287.1863	Valid
con	811	0.34777	282.0416	Valid
ma	976	0.282489	275.7094	Invalid
pr	1208	0.226006	273.0148	Invalid
qu	280	0.962199	269.4158	Invalid
over	274	0.982079	269.0896	Valid
ove	279	0.920792	256.901	Invalid
ca	1199	0.194107	232.7345	Invalid
no	593	0.379156	224.8395	Invalid
non	360	0.607083	218.5497	Valid
tr	783	0.245147	191.9502	Invalid
inte	274	0.674877	184.9163	Invalid
imp	280	0.646651	181.0624	Footprint
pa	966	0.18073	174.5849	Invalid
d	3690	0.046554	171.7838	Invalid
inter	216	0.788321	170.2774	Valid
ba	750	0.211864	158.8983	Invalid
b	3540	0.044661	158.1015	Invalid
trans	170	0.899471	152.9101	Valid
m	3455	0.043589	150.6002	Invalid
tra	343	0.438059	150.2542	Invalid
per	340	0.438144	148.9691	Valid
ha	605	0.245635	148.6094	Invalid
st	1002	0.145112	145.4025	Invalid
unde	197	0.724265	142.6802	Invalid
out	146	0.935897	136.641	Valid
pre	406	0.336093	136.4536	Valid
for	256	0.532225	136.2495	Valid
la	522	0.257016	134.1625	Invalid

Prefix	f_c	$\frac{f_c}{f_p}$	$\frac{f_c^2}{f_p}$	Semantic validity
me	680	0.196816	133.835	Invalid
hyp	226	0.582474	131.6392	Invalid
he	566	0.229801	130.0674	Invalid
t	3194	0.040296	128.7062	Invalid
gr	496	0.256331	127.1401	Invalid
mi	660	0.191027	126.0782	Invalid
an	774	0.161992	125.3822	Abbreviated
mo	656	0.18987	124.5546	Invalid
super	124	0.992	123.008	Valid
ex	564	0.216341	122.0161	Valid
ho	534	0.216809	115.7759	Invalid
pe	776	0.145182	112.6616	Invalid
pla	214	0.523227	111.9707	Invalid
li	469	0.230921	108.3018	Invalid
ch	816	0.132103	107.796	Invalid
ne	410	0.262148	107.4808	Abbreviated
under	144	0.730965	105.2589	Valid
tran	189	0.55102	104.1429	Invalid
vi	331	0.31315	103.6528	Invalid
su	846	0.12252	103.6519	Invalid
r	2847	0.035918	102.2597	Invalid
en	516	0.197929	102.1312	Valid
hyper	103	0.980952	101.0381	Valid
anti	161	0.612167	98.55894	Valid
int	406	0.239105	97.07656	Invalid
fo	481	0.197212	94.85896	Invalid
gra	215	0.433468	93.19556	Invalid
par	300	0.310559	93.1677	Valid
count	105	0.882353	92.64706	Invalid
te	539	0.168754	90.95836	Invalid
hydr	94	0.959184	90.16327	Abbreviated
wa	338	0.265515	89.74391	Invalid
ant	263	0.339793	89.36564	Abbreviated
i	2658	0.033534	89.13319	Invalid
unre	111	0.792857	88.00715	Double
po	682	0.127596	87.02039	Invalid
squ	86	1	86	Invalid
e	2607	0.032891	85.74554	Valid
aut	140	0.59322	83.05085	Abbreviated
micro	84	0.988235	83.01177	Valid
u	2561	0.03231	82.74632	Invalid
epi	110	0.743243	81.75675	Valid
coun	119	0.683908	81.38506	Invalid
counter	84	0.965517	81.10345	Valid
be	534	0.150847	80.55254	Valid
supe	125	0.64433	80.54124	Invalid
ra	474	0.166491	78.91676	Invalid

Prefix	f_c	$\frac{f_c}{f_p}$	$\frac{f_c^2}{f_p}$	Semantic validity
micr	85	0.923913	78.53261	Invalid
comp	171	0.458445	78.3941	Footprint
se	727	0.105286	76.54294	Valid
h	2463	0.031074	76.53469	Invalid
cha	249	0.305147	75.98161	Invalid
ve	282	0.266793	75.23557	Invalid
f	2439	0.030771	75.05042	Invalid
app	157	0.475758	74.69394	Footprint
auto	101	0.721429	72.86429	Valid
le	383	0.188577	72.22501	Invalid
counte	87	0.828571	72.08572	Invalid
bo	505	0.142655	72.04096	Invalid
va	274	0.259224	71.02744	Invalid

Appendix 17

Candidate suffixes

First 100 sorted on heuristic $\frac{f_c^2}{f_p}$

Suffix	f_c	$\frac{f_c}{f_p}$	$\frac{f_c^2}{f_p}$
er	4096	0.722271	2958.423
e	14375	0.181358	2607.025
ng	2819	0.892089	2514.798
ing	2654	0.941469	2498.658
ess	2494	0.938653	2341
ed	3375	0.608656	2054.216
ic	2127	0.934945	1988.628
ion	2434	0.718206	1748.113
tion	2062	0.847165	1746.855
on	3389	0.479689	1625.665
ness	2008	0.805132	1616.706
ly	3284	0.391512	1285.724
ation	1612	0.781765	1260.206
al	2194	0.571057	1252.898
y	8388	0.105825	887.6594
ss	2657	0.325214	864.0942
s	8170	0.103075	842.1193
ate	1309	0.618328	809.3911
idae	759	0.997372	757.0053
ity	951	0.793161	754.2961
ism	768	0.954037	732.7006

Suffix	f_c	$\frac{f_c}{f_p}$	$\frac{f_c^2}{f_p}$
able	895	0.774892	693.5281
us	2362	0.289107	682.8695
n	7065	0.089134	629.7292
ble	1155	0.514248	593.9559
ive	718	0.814059	584.4943
ent	926	0.620643	574.7158
ally	651	0.788136	513.0763
ist	745	0.655233	488.1487
ia	1521	0.315822	480.3657
ize	525	0.895904	470.3498
ical	497	0.911927	453.2275
dae	761	0.591298	449.9775
ceae	450	0.980392	441.1765
nt	1492	0.290725	433.7615
aceae	436	0.968889	422.4356
an	1698	0.24034	408.0968
r	5671	0.071547	405.7409
ous	968	0.409822	396.7079
d	5545	0.069957	387.9115
tive	527	0.733983	386.8092
nce	553	0.643023	355.5919
ine	684	0.517007	353.6327
le	2246	0.156243	350.9229
tic	850	0.399624	339.6803
t	5132	0.064746	332.2789
ically	325	0.970149	315.2985
te	2117	0.14727	311.7697
um	874	0.351286	307.0241
ish	425	0.711893	302.5544
a	4816	0.06076	292.619
ously	293	0.996599	292.0034
ise	602	0.453997	273.3062
ngly	280	0.965517	270.3448
ingly	274	0.978571	268.1286
sis	546	0.481482	262.8889
tor	423	0.619327	261.9751
sm	805	0.323553	260.4602
st	1137	0.221551	251.9035
ousness	239	1	239
lity	476	0.500526	238.2503
bility	268	0.884488	237.0429
usly	294	0.792453	232.9811
logy	240	0.967742	232.2581
ium	450	0.514874	231.6934
ization	226	1	226
ck	513	0.43734	224.3555
ment	454	0.490281	222.5875

Suffix	f_c	$\frac{f_c}{f_p}$	$\frac{f_c^2}{f_p}$
ology	231	0.9625	222.3375
ian	604	0.355713	214.8504
lly	826	0.251523	207.7576
sh	597	0.34669	206.9739
isation	223	0.925311	206.3444
ful	243	0.84083	204.3218
ard	296	0.671202	198.6757
ility	303	0.636555	192.876
like	214	0.887967	190.0249
ogy	248	0.765432	189.8272
l	3842	0.048472	186.2277
ics	181	0.989071	179.0219
ted	774	0.229333	177.504
cally	335	0.514593	172.3886
ter	840	0.205078	172.2656
ty	1199	0.142942	171.3878
tory	207	0.821429	170.0357
ry	1182	0.140916	166.5622
age	293	0.560229	164.1472
eae	459	0.356643	163.6993
ively	165	0.964912	159.2105
is	1134	0.1388	157.3998
ship	155	0.95092	147.3926
ated	333	0.430233	143.2674
ike	241	0.593596	143.0566
ator	245	0.579196	141.9031
ence	280	0.506329	141.7722
ative	270	0.512334	138.3302
ght	147	0.93038	136.7658
cal	545	0.248405	135.3806
ncy	201	0.672241	135.1204
ably	185	0.72549	134.2157

Appendix 18

Properties of encoded lexical relations

Primary relations

<table>
<tr><th rowspan="2">Phenomenon</th><th colspan="6">Primary relation</th></tr>
<tr><th>Lexical relation class</th><th>Relation Type</th><th>Source Lexical Record class</th><th>Encapsulating object</th><th>Source</th><th>Target</th></tr>
<tr><td>Multi-word expression with discovered component POSes</td><td>POS Specific</td><td rowspan="4">ROOT</td><td rowspan="15">POS Specific</td><td rowspan="11">Lexical Record</td><td rowspan="3">multiword expression</td><td rowspan="3">component word</td></tr>
<tr><td>Multi-word expression without discovered component POSes</td><td rowspan="3">POS Sourced</td></tr>
<tr><td>Hyphenation</td><td>hyphenation</td></tr>
<tr><td>Concatenation</td><td>concatenation</td></tr>
<tr><td>Antonymous Prefixation</td><td rowspan="11">POS Specific</td><td>ANTONYM</td><td>prefixation</td><td>unprefixed equivalent</td></tr>
<tr><td>Homonym</td><td rowspan="3">determined by morphological rule</td><td rowspan="3">derivative POS Tagged Morpheme</td><td>root POS Tagged Suffixation</td></tr>
<tr><td>Suffixation</td><td>root POS Tagged Stem</td></tr>
<tr><td></td><td>root POS Tagged Suffixation</td></tr>
<tr><td>Non-antonymous Prefixation</td><td rowspan="6">ROOT</td><td rowspan="2">prefixation</td><td>stem</td></tr>
<tr><td></td><td>prefix meaning</td></tr>
<tr><td>Redundant Stem</td><td rowspan="10">POS Tagged Stem</td><td rowspan="10">stem</td><td>alternative POS</td></tr>
<tr><td>Interpreted Stem</td><td>stem meaning</td></tr>
<tr><td rowspan="8">Analysed Stem</td><td>stem component word</td></tr>
<tr><td>stem component stem</td></tr>
<tr><td>stem component prefix meaning</td></tr>
<tr><td>stem component POS Tagged Suffixation</td></tr>
<tr><td rowspan="2">determined by morphological rule</td><td>stem component POS Tagged Stem</td></tr>
<tr><td></td></tr>
</table>

Converse relations

Phenomenon	Converse relation				Translating?
	Lexical relation class	**Relation Type**	**Source Lexical Record class**	**Encapsulating object**	
Multi-word expression with discovered component POSes	POS Specific	DERIVATIVE	POS Specific	General Lexical Record	no
Multi-word expression without discovered component POSes	POS Targeted		General	Lexicon	
Hyphenation				Lexicon	
Concatenation				Lexicon	
Antonymous Prefixation	POS Specific	ANTONYM	POS Specific	General Lexical Record	
Homonym		determined by morphological rule		POS Tagged Stem	
Suffixation				General Lexical Record	
Prefixation		DERIVATIVE		POS Tagged Stem	
Redundant Stem				General Lexical Record	yes
Interpreted Stem					no
Analysed Stem					yes
				POS Tagged Stem	no
				General Lexical Record	yes
		determined by morphological rule		POS Tagged Stem	no

Appendix 19

Formats of output files for morphological analysis

File name	Sampling rate	Column 1	Column 2	Column 3
X1Rejected concatenation components.csv				
X1Concatenations with components.csv		the word analysed		
likewise X2,X3				
WordsWithAntonymousPrefixes.csv		antonymous prefixation	unprefixed equivalent (candidate antonym)	
Primary Identical words Results.csv				
Primary Identical words Result Samples.csv	1/100	derivative	derivative POS	root
Primary Monosyllabic Identical words .csv		derivative backwards	derivative	derivative POS
Suffixes.csv		suffix		
Prefixes.csv		prefix	f_c	f_c / f_p
X1 Suffix-stripping Results.csv				
X1 Suffix-stripping Result Samples.csv	1/100	derivative	derivative POS	root
X1 monosyllabic roots.csv		derivative backwards	derivative	derivative POS
likewise X2, X3, X4, X5, X6				
X1 unidentified roots.csv		word with no root identified backwards	word with no root identified	POS of word with no root identified
likewise X2, X3, X4, X5, X6				
Irregular rejected prefixation components.csv		Word rejected as an irregular prefixation		
Irregular prefixations with components.csv		Word accepted as an irregular prefixation	prefix name	
X1Prefixations with components.csv				
X1Residual antonymous prefixes.csv		antonymous prefixation	unprefixed equivalent (candidate antonym)	
likewise X2, X3, X4, X5, X6, X7, X8				
Residual antonymous prefixes.csv		antonymous prefixation	unprefixed equivalent (candidate antonym)	
Stem relations from stem dictionary pruning.csv		alternative word	alternative POS	stem
Affixation stems1.csv				
Affixation stems summary1.csv	1/100			
Affixation stems2.csv			number of prefixes	number of suffixes
Affixation stems summary2.csv	1/100	stem		

File name	Sampling rate	Column 1	Column 2	Column 3
StemsX0components.csv		stem	"Prefix:"	
likewise X1, X2, X3, X4				
StemsX0 Lexical restorations.csv		stem	stem POS	prefix
likewise X1, X2, X3, X4				

File name	Column 4	Column 5	Column 6	Column 7
Primary Identical words Results.csv			devative suffix POS	
Primary Identical words Result Samples.csv	root POS	derivative suffix		root suffix
Primary Monosyllabic Identical words .csv	root	root POS	derivative suffix	devative suffix POS
Suffixes.csv				
Prefixes.csv	f_c^2 / f_p	q_s	d	f_p
X1 Suffix-stripping Results.csv			derivative suffix POS	
X1 Suffix-stripping Result Samples.csv	root POS	derivative suffix		root suffix
X1 monosyllabic roots.csv	root	root POS	derivative suffix	devative suffix POS
Irregular prefixations with components.csv				
X1Prefixations with components.csv			stem	
Stem relations from stem dictionary pruning.csv	stem POS	relation type		
Affixation stems1.csv				
Affixation stems summary1.csv				
Affixation stems2.csv				
Affixation stems summary2.csv	"Prefixes:"			
StemsX0 Lexical restorations.csv	"Suffix:"	suffix		
likewise X1, X2, X3, X4				

File name	Column 8	Column 9	Remainder
X1Rejected concatenation components.csv			rejected components
X1Concatenations with components.csv			up to 5 accepted components arranged in so that if there is are 3 components, they occupy columns 2, 4 & 6
likewise X2,X3			
Primary Identical words Results.csv	root suffix POS		
Primary Identical words Result Samples.csv			
Primary Monosyllabic Identical words .csv	root suffix	root suffix POS	
Suffixes.csv	f_c - f_d		
Prefixes.csv			
X1 Suffix-stripping Results.csv	root suffix POS		
X1 Suffix-stripping Result Samples.csv			
X1 monosyllabic roots.csv	root suffix	root suffix POS	
likewise X2, X3, X4, X5, X6			
Affixation stems1.csv			an indefinite number of prefixes, followed by "Suffixes:", followed by an indefinite number of suffixes
Affixation stems summary1.csv			
Affixation stems2.csv			
Affixation stems summary2.csv			
StemsX0components.csv			
likewise X1, X2, X3, X4			
StemsX0 Lexical restorations.csv			
likewise X1, X2, X3, X4			

Appendix 20

Formats of input files for morphological analysis

File name	Column 1	Column 2	Column 3	Column 4	Remaining columns
Suffix stripping stoplist.csv	false derivative word	false derivative POS	false root word	false root POS	
Secondary suffix stripping stoplist.csv					
Irregular prefixes.csv	footprint	prefix name	character sequence to be deleted	character sequence to be inserted	instances
Detailed Prefix meanings.csv	prefix name	meaning	meaning POS		meaning and meaning POS an indefinite number of times
Detailed Irregular prefix meanings.csv					
Prefixation stem stoplist.csv	false stem	false stem POS			
Linking vowel exceptions.csv	prefix with superfluous linking vowel	stem with missing initial vowel			
Reverse linking vowel exceptions.csv	prefix without linking vowel	stem with superfluous initial vowel			
Final suffixation reprieves.csv	word reprieved	POS of word reprieved			
Stem meanings.csv	stem	stem POS	stem meaning	stem meaning POS	3 pairs of columns, each pair containing stem meaning followed by stem meaning POS
					an indefinite number of associated prefixes
					#
					an indefinite number of associated

File name	Column 1	Column 2	Column 3	Column 4	Remaining columns
					suffixes
Lexical restoration stoplist.csv	tem homonym	stem homonym POS			

Appendix 21

Suffixation Analysis Algorithm

```
for each word in the atomic dictionary
{
  create Map<POSTaggedMorpheme, POSTaggedSuffixation>;
  for each POS of the current word
  {
    create POSTaggedWord from current word / POS;
    while the Map is empty and there are untried suffixes in the
    secondary suffix set
    {
      get next pre-identified suffix from secondary suffix set
      if current word ends with current pre-identified suffix
      {
        POSTaggedSuffixation is result of applying root
        identification algorithm to the POSTaggedWord using the
        current pre-identified suffix (§5.2.2);
        if the POSTaggedSuffixation is valid
        {
          add to the Map a mapping from current word as a
          POSTaggedMorpheme to the POSTaggedSuffixation;
        }
      }
      if Map is empty
      {
        write POSTaggedWord to unidentified roots file;
      }
    }
    for each entry in the Map
    {
      if POSTaggedSuffixation is monosyllabic and the rule which
      generated is inapplicable to monosyllables
      {
        reject entry;
      }
      else if POSTaggedSuffixation's Relation.Type is DERIV
      {
        reject entry;
      }
      else
      {
        remove the POSTaggedMorpheme from the atomic dictionary;
        encode LexicalRelation of POSTaggedSuffixation's Type between
        POSTaggedMorpheme and POSTaggedSuffixation;
      }
    }
  }
}
```

Appendix 22

Relation types with their converses

Relation types in **bold** exist in Princeton WordNet. All their converses have been implemented in the model of WordNet described in this thesis. Types not in bold, whose converses are also not in bold have been implemented for lexical relations only. The five types which are their own converses appear at the bottom of the table. Each relation type represents a semantic or syntactic transformation, or a combination of a syntactic transformation with one or more semantic transformations. Relations whose type category is "WordNet" are never used in the morphological analysis, some having been eliminated from the model (§4.3). Relations whose type category is "Derivational" specify only the direction of derivation, except for type DERIV which specifies only that a morphological relationship exists[20]. Each lexical link is the combination of two relations which are converses of each other. Type **SYNONYM** is redundant except for lexical relations.

Relation type	Converse Relation Type	Relation Type Category	Lexical Links
HYPERNYM	**HYPONYM**	Semantic	0
ENTAILMENT	COUNTER_ENTAILMENT	Semantic	0
CAUSE	EFFECT	Semantic	484
INSTANCE	INSTANTIATED	WordNet	0
SIMILAR	**CLUSTERHEAD**	WordNet	0
MEMBER_MERONYM	**MEMBER_HOLONYM**	WordNet	0
SUBSTANCE_MERONYM	**SUBSTANCE_HOLONYM**	Semantic	2348
PART_MERONYM	**PART_HOLONYM**	Semantic	0
ATTRIBUTE	**ATTRIBUTE_VALUE**	Semantic	4791
CLASS_MEMBER	MEMBER_CLASS	WordNet	0
SEE_ALSO	SEEN_ALREADY	WordNet	0
PARTICIPLE	VERB_SOURCE	Syntactic	3778
PERTAINYM	PERTAINER	Semantic	6646
ROOT	DERIVATIVE	Derivational	174052
ANTONYM_OF_ATTRIBUTE_VALUE	ATTRIBUTE_OF_ANTONYM	Semantic	319
ANTONYM_OF_PARTICIPLE	VERBSOURCE_OF_ANTONYM	Semantic / Syntactic	8
GERUND	VERBSOURCE_OF_GERUND	Syntactic	4299
MEASUREDBY	MEASURING	Semantic	65
PATIENT	AFFECTING	Semantic	146
ABLE	POTENTIAL	Semantic	574
QUALIFIED	QUALIFYING	Semantic	927
RESEMBLING	RESEMBLEDBY	Semantic	173
DEMONSTRATE	DEMONSTRATION	Semantic	5
SUBJECT	ROLE	Syntactic	3118
POSSESSION_OF_ATTRIBUTE	POSSESSOR_OF_ATTRIBUTE	Semantic	318
SUBJECT_OF_VERBSOURCE_OF_GERUND	GERUND_OF_ROLE	Syntactic	97
BELIEVE_PRACTICE	OBJECT_OF_BELIEF_PRACTICE	Semantic	107
GERUND_OF_BELIEVE_PRACTICE	OBJECT_OF_BELIEF_PRACTICE_OF_ VERBSOURCE_OF_GERUND	Semantic / Syntactic	562
GERUND_OF_BELIEVE_PRACTICE_PERTAINYM	PERTAINER_TO_OBJECT_OF_BELIEF_PRACTICE_OF_ VERBSOURCE_OF_GERUND	Semantic / Syntactic	170
SUBJECT_OF_BELIEVE_PRACTICE	OBJECT_OF_BELIEF_PRACTICE_OF_ROLE	Semantic / Syntactic	659
SUBJECT_OF_BELIEVE_PRACTICE_ PERTAINYM	PERTAINER_TO_OBJECT_OF_BELIEF_PRACTICE_OF_ ROLE	Semantic / Syntactic	135
SINGULAR	PLURAL	Semantic	2608
MASCULINE	FEMININE	Semantic	228
DESTINATION	DIRECTION	Semantic	7
COMPARISON	ADJECTIVE_SOURCE	Syntactic	49

[20] All lexical relations have a supertype which specifies the direction of derivation. Only the DERIV relations between WordNet word senses do not provide this information.

Relation type	Converse Relation Type	Relation Type Category	Lexical Links
HOME	INHABITANT	Semantic	820
FULLSIZE	DIMINUTIVE	Semantic	1604
REPEATED	REPETITION	Semantic	116
AFFECTED_ORGAN	DISEASE	Semantic	105
ABILITY	POTENTIALITY	Semantic	11
ANTONYM	**ANTONYM**	Semantic	3444
DERIV	**DERIV**	Derivational	4820
SYNONYM	**SYNONYM**	Semantic	750
VERB_GROUP_POINTER	**VERB_GROUP_POINTER**	WordNet	0
NEARSYNONYM	NEARSYNONYM	Semantic	459
		TOTAL	218802

Appendix 23

Preposition taxonomy by digraph analysis

(after Litkowski, 2002)

Primitive?	Strong components
n	over, above
n	against
n	but
n	along
n	on
n	via, by way of
n	through
n	touching
n	until, up to
n	below, underneath
n	inside, within
n	in favour of, along with, with respect to, in proportion to, in relation to, in connection with, with reference to, in respect of, as regards, concerning, about, with, in place of, instead of, in support of, except, other than, apart from, in addition to, behind, beside, next to, following, past, beyond, after, to, before, in front of, ahead of, for, by, according to
y	in
n	across
n	by means of
n	in the course of
n	during
n	on behalf of
y	of
y	than
y	as
y	from
y	by reason of, because of, on account of
y	as far as
y	including

Appendix 24

Preposition record fields

Type	Name	XML element	
String	wordForm;	<hw>	
short	WordnetSenseNumber;		obtained by counting <S> elements
String	register;	<reg>	
short	tppSenseNumber;	<b>	0 if none
String	tppSenseid;	<senseid>	0 if none
String	geography;	<ge>	
String	gloss;	<df>	
String[]	adjectiveExamples;	<eg>, <ex>, <gg>	an indefinite number, as determined by <gg> elements
String[]	conjunctionExamples;	<eg>, <ex>, <gg>	an indefinite number, as determined by <gg> elements
String[]	adverbExamples;	<eg>, <ex>, <gg>	an indefinite number, as determined by <gg> elements
String[]	examples;	<eg>, <ex>, <gg>	preposition examples: an indefinite number, as determined by <gg> elements
String	superordinateTaxonomicCategorizer;	<sup>	converted to uppercase
String	semanticRoleType;	<srtype>	converted to uppercase
List<String>	synonyms;	<opreeps>	parentheses and numerals removed
String	complementProperties;	<cprop>	converted to uppercase
String	relationToCoreSense;	<srel>	converted to uppercase
Boolean	currentSynonymMatched;		used in synonym identification
Boolean	currentSynonymMatchAccepted;		used in synonym identification
Boolean	currentSynonymMatchReinforced;		used in synonym identification
Boolean	currentValidSynonym;		used in synonym identification
List<PrepositionRecord>	validSynonyms;		additional synonyms identified by variant spellings and from synonym identification
Boolean	currentValidHypernym;		
List<PrepositionRecord>	validHypernyms;		hypernyms identified among multiple synonym senses during synonym identification
List<PrepositionRecord>	validHyponyms;		hyponyms identified among multiple synonym senses during synonym identification
Preposition	instance;		the Preposition created from this Preposition record
int	synsetID;		the ID of the Preposition and of the Synset to which the Preposition is assigned

Appendix 25

Superordinate taxonomic categorizers

ACTIVITY
AGENT
BACKDROP
BARRIER
CAUSE
CONSEQUENCE
DOUBLES
DOUBLES; SCALAR
EXCEPTION
MEANSMEDIUM
MEMBERSHIP
PARTY
POSSESSION
QUANTITY
SCALAR
SCALAR; TEMPORAL
SPATIAL
SPATIAL; TEMPORAL
SUBSTANCE
TANDEM
TARGET
TEMPORAL
TOPIC
TRIBUTARY
VOID

Appendix 26

Top ontology for prepositions

Word forms	Hypernym wordforms
à la:	like:
a cut above:	above:
abaft:	behind:
aboard:onto:on:	on:onto:
about:	with reference to
about:around:round:	around:round:
above:	above:o'er:over:
above:o'er:over:	not at
above:on top of:over:atop:o'er:	above:o'er:over:
absent:minus:	sans:without:
according to:	with reference to
according to:depending on:	according to:
across:	via
across:opposite:	across:
afore:before:fore:	not at
afore:before:fore:in front of:	afore:before:fore:

Word forms	Hypernym wordforms
afore:before:fore:previous to:	afore:before:fore:
after the fashion of:	like:
after:	past:
after:subsequent to:	after:
against:agin:	with:
against:agin:up against:	against:agin:
against:agin:versus:	against:agin:
against:agin:with:	against:agin:
ahead of:	afore:before:fore:
ahead of:in front of:	afore:before:fore:
all for:	for:
all over:	thro':through:thru:throughout:up and down:
along with:	with:
along:	via
alongside:	along:
alongside:by:	along:
amid:amidst:	mongst:among:amongst:
anent:	about:
anti:	against:agin:
apart from:	sans:without:
apropos:as for:	about:
around:round:	not at
as far as:	to:
as from:	frae:from:
as of:	frae:from:
as regards:	about:
as to:	about:
as well as:	apart from:
as:qua:	as
aside from:	apart from:
aslant:	across:
astraddle:	on:onto:
astride:	on:onto:
at a range of:	at:
at the hand of:	by:
at the hands of:	by:
at the heels of:	behind:
athwart:thwart:	afore:before:fore:
back of:	behind:
bar:	apart from:
bare of:	apart from:
barring:	sans:without:
because of:on account of:by reason of:owing to:	due to:
behind:	past:
behind:beneath:underneath:neath:under:	behind:
behind:in back of:	behind:
below:beneath:under:neath:	beneath:neath:
below:under:	beneath:neath:
below:under:underneath:beneath:neath:	beneath:neath:
beneath:neath:	not at
beside:	with:
beside:besides:in addition to:on top of:	apart from:

Word forms	Hypernym wordforms
beside:next to:	near:nigh:
between:betwixt:	among:between
beyond:	beyond:past:
beyond:past:	not at
but:except for:with the exception of:excepting:save:but for:except:	apart from:
by courtesy of:courtesy of:	due to:
by dint of:	by:
by force of:by means of:by way of:	by:
by the hand of:	by:
by the hands of:	by:
by the name of:	as
by virtue of:	due to:
by way of:	as
by way of:through:via:thro':thru:	via
by:	caused by
by:on the part of:	by:
care of:	chez:
cept:	apart from:
circa:	around:round:
come:	at:
complete with:	with:
concerning:on:over:in connection with:o'er:	about:
considering:given:	despite:in spite of:notwithstanding:for all:in the face of:
contrary to:	against:agin:
counting:	with:
cum:	with:
dehors:	outside:outwith:
despite:in spite of:notwithstanding:for all:in the face of:	not caused by
down:	via
down:throughout:	thro':through:thru:throughout:up and down:
due to:	caused by
during:in the course of:	in:
during:in:in the course of:	in:
ere:	afore:before:fore:
ex:	out of:outta:
excluding:exclusive of:	apart from:
failing:	sans:without:
following:	after:
for the benefit of:	for:
for:	as why
for:on behalf of:	for:
forbye:	apart from:
fornent:	near:nigh:
frae:from:	away from
frae:from:	by:
frae:from:	at:
gainst:	against:agin:
give or take:	as not

Word forms	Hypernym wordforms
gone:	after:
having regard to:	about:
in accord with:	according to:
in advance of:	afore:before:fore:
in aid of:	for:
in bed with:	with:
in behalf of:	for:
in behalf of:on behalf of:	for:
in case of:	against:agin:
in common with:	like:
in company with:	with:
in consideration of:	due to:
in contravention of:	against:agin:
in default of:	sans:without:
in excess of:over:upward of:upwards of:o'er:	above:o'er:over:
in face of:	afore:before:fore:
in favor of:	for:
in favour of:	for:
in front of:	afore:before:fore:
in honor of:	for:
in honour of:	for:
in keeping with:	according to:
in lieu of:instead of:in place of:	as not
in light of:	considering:given:
in line with:	according to:
in memoriam:	for:
in need of:	sans:without:
in peril of:	against:agin:
in peril of:	afore:before:fore:
in proportion to:	according to:
in proportion to:in relation to:	according to:
in re:	in case of:
in reference to:	with reference to
in regard to:	about:
in respect of:	with reference to
in sight of:	near:nigh:
in terms of:	with reference to
in the face of:	afore:before:fore:
in the fashion of:	like:
in the grip of:in the teeth of:	against:agin:
in the light of:	with reference to
in the matter of:	with reference to
in the midst of:under:	mongst:among:amongst:
in the name of:	for:
in the pay of:	for:
in the person of:	as
in the shape of:	as
in the teeth of:	against:agin:
in the throes of:	mongst:among:amongst:
in token of:	due to:
in view of:	due to:
in virtue of:	due to:

Word forms	Hypernym wordforms
in:	at:
in:inside:	in:
in:under:	in:
including:	with:
inclusive of:	with:
inside of:	in:
inside:	in:
into:	to:
irrespective of:	apart from:
less:minus:	sans:without:
like:	with reference to
like:on the order of:	like:
little short of:	near:nigh:
mid:	mongst:among:amongst:
midst:	mongst:among:amongst:
minus:	sans:without:
mod:	apart from:
modulo:	apart from:
mongst:among:amongst:	among:between
mongst:among:amongst:between:betwixt:	mongst:among:amongst:
more like:	near:nigh:
near to:	near:nigh:
near:nigh:	near:with
next door to:	near:nigh:
next to:	near:nigh:
nothing short of:	near:nigh:
o':of:	with reference to
o'er:over:	above:o'er:over:
o'er:over:on top of:	above:o'er:over:
o'er:over:via:	by:
of the name of:	as
of the order of:	around:round:
of the order of:on the order of:	around:round:
off:	beyond:past:
off:	frae:from:
on a level with:	near:nigh:
on a level with:on a par with:	near:nigh:
on pain of:under pain of:	under:
on the point of:	afore:before:fore:
on the score of:	due to:
on the strength of:	due to:
on the stroke of:	at:
on top of:	on:onto
on:	at:
on:	above:o'er:over:
opposite:	afore:before:fore:
other than:	apart from:
out of keeping with:	regardless of:
out of line with:	regardless of:
out of:outta:	frae:from:
outboard of:	outside:outwith:
outside of:	outside:outwith:
outside:outwith:	not at

Word forms	Hypernym wordforms
over against:	against:agin:
over and above:	apart from:
overtop:	above:o'er:over:
pace:	for:
pace:	against:agin:
past:	beyond:past:
pending:	afore:before:fore:
per:	in:
plus:	with:
pon:upon:on:	on:
preparatory to:	for:
prior to:	afore:before:fore:
pro:	for:
pursuant to:under:	according to:
re:	about:
regarding:	about:
regardless of:	with reference to
relative to:	with reference to
respecting:	with reference to
round about:	around:round:
round:	around:round:
sans:without:	give or take:
saving:	apart from:
short for:	in lieu of:instead of:in place of:
short of:	apart from:
since:	after:
than:	with reference to
than:	as not
thanks to:	due to:
this side of:	afore:before:fore:
thro':through:thru:	via
thro':through:thru:throughout:up and down:	at:
till:until:while:	afore:before:fore:
to the accompaniment of:	with:
to the tune of:	as
to:	toward:towards:
to:	for:
to:	at:
together with:	with:
touching:	about:
toward:towards:	with reference to
toward:towards:	not at
under cover of:	under:
under sentence of:	under:
under the heel of:	under:
under:	beneath:neath:
under:underneath:	beneath:neath:
unlike:	with reference to
unto:	to:
up against:	against:agin:
up and down:	along:
up before:	afore:before:fore:

Word forms	Hypernym wordforms
up for:	afore:before:fore:
up to:	at:
up:	via
upside:	against:agin:
versus:	against:agin:
via:	by:
vice:	in lieu of:instead of:in place of:
vis-à-vis:	about:
with regard to:	with reference to
with respect to:	with reference to
withal:	with:
within sight of:	near:nigh:
within:	in:
on:onto	on:
on:onto	to:
away from	with reference to
away from	not at
via	at:
via	not at
chez	at:
among:between	with:
with:	give or take:
with:	near:with
caused by	as why
not caused by	as not why
as why	as
as not why	as not

Appendix 27

Preposition antonyms

Word forms	Antonym wordforms
above:o'er:over:	beneath:neath:
according to:	regardless of:
across:	along:
afore:before:fore:	beyond:past:
against:agin:	for:
along:	across:
at:	not at
beneath:neath:	above:o'er:over:
despite:in spite of:notwithstanding:for all:in the face of:	due to:
down:	up:
due to:	despite:in spite of:notwithstanding:for all:in the face of:
for:	against:agin:
frae:from:	to:
in keeping with:	out of keeping with:
in line with:	out of line with:
in:	outside:outwith:
like:	unlike:
out of keeping with:	in keeping with:
out of line with:	in line with:
outside:outwith:	in:
beyond:past:	afore:before:fore:
regardless of:	according to:
sans:without:	near:with
to:	frae:from:
toward:towards:	away from
unlike:	like:
up:	down:
near:with	sans:without:
on:onto	off:
away from	toward:towards:
not at	at:
as	as not
as not	as
caused by	not caused by
not caused by	caused by
as why	as not why
as not why	as why

Appendix 28

Adjective to adjective pertainyms

Synset ID	Word form	Synset ID	Word form	New relation type
303048385	bilabial	302754417	labial	SIMILAR
302891733	protozoological	302891444	zoological	SIMILAR
302894327	sensorineural	302894119	neural	SIMILAR
302885790	subclinical	302885529	clinical	DERIV
303080492	Latin	303080351	Romance	SIMILAR
302846743	antediluvian	302846630	diluvial	DERIV
302846743	antediluvial	302846630	diluvial	DERIV
303096747	parenteral	303096635	parenteral	DERIV
302833873	antibacterial	302833544	bacterial	DERIV
302838220	bipolar	302838005	polar	SIMILAR
302750166	intracranial	302844273	cranial	DERIV
303030096	pre-Columbian	303029984	Columbian	DERIV
303009792	fibrocalcific	303009696	calcific	SIMILAR
303014941	lumbosacral	303014770	lumbar	SIMILAR
303014941	lumbosacral	303113164	sacral	SIMILAR
303015336	biflagellate	303015113	flagellate	SIMILAR
302717021	socioeconomic	302716605	economic	SIMILAR
302991962	cross-sentential	302991690	sentential	SIMILAR
302991819	intrasentential	302991690	sentential	SIMILAR
303003031	thermohydrometric	303002841	hydrometric	SIMILAR
303003031	thermogravimetric	303002841	hydrometric	SIMILAR
302728303	bifilar	302728113	filar	SIMILAR
302728444	unifilar	302728113	filar	SIMILAR
302982956	thalamocortical	302974979	cortical	SIMILAR
302982840	cortico-hypothalamic	302982729	hypothalamic	SIMILAR
302981508	antithyroid	302981329	thyroid	DERIV
302948198	interlobular	302948068	lobular	DERIV
302948281	intralobular	302948068	lobular	DERIV
302946777	transatlantic	302946507	Atlantic	DERIV
302645868	astomatal	302645494	stomatal	ANTONYM
302649570	biauricular	302649125	auricular	SIMILAR
302933807	dizygotic	302882275	zygotic	SIMILAR
302933807	dizygous	302882275	zygotic	SIMILAR
302933692	monozygotic	302882275	zygotic	SIMILAR
302933230	intrauterine	302933132	uterine	DERIV
302936627	monomorphemic	302936410	morphemic	SIMILAR
302936764	polymorphemic	302936410	morphemic	SIMILAR
302936511	bimorphemic	302936410	morphemic	SIMILAR

Appendix 29

Exceptions specified in implementing the WordNet model.

All the following Exceptions are implemented as subclasses of **`WordnetBuilderException`**.

- **`DataFormatException`**
- **`DuplicateGlossException`**
- **`DuplicateRelationException`**
- **`DuplicateSensekeyException`**
- **`DuplicateWordNumberException`**
- **`InconsistentLexiconException`**
- **`InconsistentWordnetException`**
- **`LemmaMismatchException`**
- **`LexicalOmissionException`**
- **`MixedVerbFrameTypesException`**
- **`NonLexicalFrameException`**
- **`Paradox`**
- **`UnexpectedParseException`**
- **`UnexpectedPOSException`**
- **`UnexpectedXMLFormatException`**
- **`UnknownSynsetException`**
- **`UnmatchedFrameException`**

Appendix 30

Morphological rules for "-ion" suffix

Source		Target		
Morpheme	**POS**	**Morpheme**	**POS**	**Relation**
ce	VERB	cion	NOUN	GERUND
construct	VERB	construction	NOUN	GERUND
construe	VERB	construction	NOUN	GERUND
ct	VERB	ction	NOUN	GERUND
ct	ADJECTIVE	ction	NOUN	ATTRIBUTE
fy	VERB	faction	NOUN	GERUND
join	VERB	junction	NOUN	GERUND
suck	VERB	suction	NOUN	GERUND
uce	VERB	uction	NOUN	GERUND
here	VERB	hesion	NOUN	GERUND
her	VERB	hesion	NOUN	GERUND
ete	VERB	etion	NOUN	GERUND
ete	ADJECTIVE	etion	NOUN	ATTRIBUTE
rete	VERB	retion	NOUN	GERUND
ect	VERB	exion	NOUN	GERUND
suspect	VERB	suspicion	NOUN	GERUND
ise	ADJECTIVE	ision	NOUN	ATTRIBUTE
appear	VERB	apparition	NOUN	GERUND
define	VERB	definition	NOUN	GERUND
ise	VERB	ition	NOUN	GERUND
ize	VERB	ition	NOUN	GERUND

Source		Target		Relation
Morpheme	**POS**	**Morpheme**	**POS**	
ish	VERB	ition	NOUN	GERUND
ite	ADJECTIVE	ition	NOUN	ATTRIBUTE
nourish	VERB	nutrition	NOUN	GERUND
ose	VERB	osition	NOUN	GERUND
peat	VERB	petition	NOUN	GERUND
pete	VERB	petition	NOUN	GERUND
quire	VERB	quisition	NOUN	GERUND
render	VERB	rendition	NOUN	GERUND
l	VERB	llion	NOUN	GERUND
pel	VERB	pulsion	NOUN	GERUND
nd	VERB	nsion	NOUN	GERUND
sent	VERB	sension	NOUN	GERUND
nd	VERB	ntion	NOUN	GERUND
vene	VERB	vention	NOUN	GERUND
move	VERB	motion	NOUN	GERUND
ceive	VERB	ception	NOUN	GERUND
deem	VERB	demption	NOUN	GERUND
orb	VERB	orption	NOUN	GERUND
scribe	VERB	scription	NOUN	GERUND
ume	VERB	umption	NOUN	GERUND
merge	VERB	mersion	NOUN	GERUND
rt	VERB	rsion	NOUN	GERUND
rt	ADJECTIVE	rsion	NOUN	ATTRIBUTE
ur	VERB	ursion	NOUN	GERUND
se	VERB	sion	NOUN	GERUND
de	VERB	sion	NOUN	GERUND
cede	VERB	cession	NOUN	GERUND
ceed	VERB	cession	NOUN	GERUND
mit	VERB	mission	NOUN	GERUND
ss	VERB	ssion	NOUN	GERUND
t	VERB	tion	NOUN	GERUND
olve	VERB	olution	NOUN	GERUND
ute	ADJECTIVE	ution	NOUN	ATTRIBUTE

Appendix 31

Morphological rules for "-al" suffix

Source		Target		Relation
Morpheme	**POS**	**Morpheme**	**POS**	
ous	ADJECTIVE	al	ADJECTIVE	NEARSYNONYM
um	NOUN	al	ADJECTIVE	PERTAINER
on	NOUN	al	ADJECTIVE	PERTAINER
a	NOUN	al	ADJECTIVE	PERTAINER
us	NOUN	al	ADJECTIVE	PERTAINER
	VERB	al	NOUN	GERUND
duke	NOUN	ducal	ADJECTIVE	PERTAINER
y	NOUN	ical	ADJECTIVE	DERIVATIVE
ex	NOUN	ical	ADJECTIVE	DERIVATIVE
ix	NOUN	ical	ADJECTIVE	DERIVATIVE

Source		Target		
Morpheme	**POS**	**Morpheme**	**POS**	**Relation**
	NOUN	ical	ADJECTIVE	PERTAINER
y	NOUN	ical	ADJECTIVE	PERTAINER
ice	NOUN	ical	ADJECTIVE	PERTAINER
d	NOUN	dal	ADJECTIVE	PERTAINER
de	NOUN	dal	ADJECTIVE	PERTAINER
ea	NOUN	eal	ADJECTIVE	PERTAINER
nx	NOUN	ngeal	ADJECTIVE	PERTAINER
h	NOUN	hal	ADJECTIVE	PERTAINER
ce	NOUN	cial	ADJECTIVE	PERTAINER
cy	NOUN	cial	ADJECTIVE	PERTAINER
x	NOUN	cial	ADJECTIVE	PERTAINER
t	NOUN	cial	ADJECTIVE	PERTAINER
	NOUN	ial	ADJECTIVE	PERTAINER
nce	NOUN	ncial	ADJECTIVE	PERTAINER
or	NOUN	orial	ADJECTIVE	PERTAINER
r	NOUN	rial	ADJECTIVE	PERTAINER
ce	NOUN	tial	ADJECTIVE	PERTAINER
cy	NOUN	tial	ADJECTIVE	PERTAINER
t	NOUN	tial	ADJECTIVE	PERTAINER
verb	NOUN	verbial	ADJECTIVE	PERTAINER
m	NOUN	mal	ADJECTIVE	PERTAINER
de	NOUN	dinal	ADJECTIVE	PERTAINER
ne	NOUN	nal	ADJECTIVE	PERTAINER
n	NOUN	nal	ADJECTIVE	PERTAINER
ude	NOUN	udinal	ADJECTIVE	PERTAINER
pe	NOUN	pal	ADJECTIVE	PERTAINER
re	NOUN	ral	ADJECTIVE	PERTAINER
er	NOUN	ral	ADJECTIVE	PERTAINER
ra	NOUN	ral	ADJECTIVE	PERTAINER
or	NOUN	ral	ADJECTIVE	PERTAINER
r	NOUN	ral	ADJECTIVE	PERTAINER
pose	VERB	posal	NOUN	GERUND
se	NOUN	sal	ADJECTIVE	PERTAINER
ss	NOUN	sal	ADJECTIVE	PERTAINER
ct	NOUN	ctal	ADJECTIVE	PERTAINER
it	NOUN	ital	ADJECTIVE	PERTAINER
nt	NOUN	ntal	ADJECTIVE	PERTAINER
st	NOUN	stal	ADJECTIVE	PERTAINER
ty	NOUN	tal	ADJECTIVE	PERTAINER
t	VERB	ttal	NOUN	GERUND
	NOUN	ual	ADJECTIVE	PERTAINER
ive	NOUN	ival	ADJECTIVE	PERTAINER
ive	ADJECTIVE	ival	ADJECTIVE	NEARSYNONYM
ove	VERB	oval	NOUN	GERUND
w	VERB	wal	NOUN	GERUND

Appendix 32

Morphological rules for "-ant" suffix

Source		Target		Relation
Morpheme	**POS**	**Morpheme**	**POS**	
ate	VERB	ant	ADJECTIVE	PARTICIPLE
y	VERB	ant	ADJECTIVE	PARTICIPLE
ate	VERB	ant	NOUN	GERUND
	VERB	ant	NOUN	GERUND
ess	VERB	essant	ADJECTIVE	PARTICIPLE
y	VERB	iant	ADJECTIVE	PARTICIPLE
y	VERB	iant	NOUN	GERUND
idise	VERB	idant	ADJECTIVE	PARTICIPLE
idise	VERB	idant	NOUN	GERUND
	NOUN	inant	NOUN	DIMINUTIVE
in	VERB	inant	ADJECTIVE	PARTICIPLE
in	VERB	inant	NOUN	GERUND
ll	VERB	lant	ADJECTIVE	PARTICIPLE
ll	VERB	lant	NOUN	GERUND
nd	VERB	ndant	ADJECTIVE	PARTICIPLE
nd	VERB	ndant	NOUN	GERUND
er	VERB	rant	ADJECTIVE	PARTICIPLE
re	VERB	rant	ADJECTIVE	PARTICIPLE
er	VERB	rant	NOUN	GERUND
re	VERB	rant	NOUN	GERUND
rd	VERB	rdant	ADJECTIVE	PARTICIPLE
rd	VERB	rdant	NOUN	GERUND
se	VERB	sant	ADJECTIVE	PARTICIPLE
se	VERB	sant	NOUN	GERUND
t	VERB	tant	ADJECTIVE	PARTICIPLE
te	VERB	tant	ADJECTIVE	PARTICIPLE
t	VERB	tant	NOUN	GERUND
te	VERB	tant	NOUN	GERUND
ue	VERB	uant	ADJECTIVE	PARTICIPLE
ue	VERB	uant	NOUN	GERUND
ounce	VERB	unciant	ADJECTIVE	PARTICIPLE
ounce	VERB	unciant	NOUN	GERUND
ound	VERB	undant	NOUN	GERUND
ve	VERB	vant	ADJECTIVE	PARTICIPLE
ve	VERB	vant	NOUN	GERUND

Appendix 33

Morphological rules for "-ent" suffix

Source		Target		Relation
Morpheme	**POS**	**Morpheme**	**POS**	
b	VERB	bent	ADJECTIVE	PARTICIPLE
b	VERB	bent	NOUN	GERUND
de	VERB	dent	ADJECTIVE	PARTICIPLE
de	VERB	dent	NOUN	GERUND
dge	VERB	dgment	NOUN	GERUND
er	VERB	erent	ADJECTIVE	PARTICIPLE
ere	VERB	erent	ADJECTIVE	PARTICIPLE
er	VERB	erent	NOUN	GERUND
ere	VERB	erent	NOUN	GERUND
ge	VERB	gent	ADJECTIVE	PARTICIPLE
ge	VERB	gent	NOUN	GERUND
ain	VERB	inent	ADJECTIVE	PARTICIPLE
ain	VERB	inent	NOUN	GERUND
ist	VERB	istent	ADJECTIVE	PARTICIPLE
ist	VERB	istent	NOUN	GERUND
itt	VERB	ittent	ADJECTIVE	PARTICIPLE
itt	VERB	ittent	NOUN	GERUND
ll	VERB	lent	ADJECTIVE	PARTICIPLE
ll	VERB	lent	NOUN	GERUND
l	VERB	llent	ADJECTIVE	PARTICIPLE
l	VERB	llent	NOUN	GERUND
	VERB	ment	NOUN	DERIVATIVE
er	VERB	ment	NOUN	DERIVATIVE
nd	VERB	ndent	ADJECTIVE	PARTICIPLE
nd	VERB	ndent	NOUN	GERUND
neglect	VERB	negligent	ADJECTIVE	PARTICIPLE
obey	VERB	obedient	ADJECTIVE	PARTICIPLE
ound	VERB	onent	ADJECTIVE	PARTICIPLE
ose	VERB	onent	ADJECTIVE	PARTICIPLE
ound	VERB	onent	NOUN	GERUND
ose	VERB	onent	NOUN	GERUND
rr	VERB	rrent	ADJECTIVE	PARTICIPLE
r	VERB	rrent	ADJECTIVE	PARTICIPLE
rr	VERB	rrent	NOUN	GERUND
r	VERB	rrent	NOUN	GERUND
sce	VERB	scent	ADJECTIVE	PARTICIPLE
sce	VERB	scent	NOUN	GERUND
sense	VERB	sentient	ADJECTIVE	PARTICIPLE
sense	VERB	sentient	NOUN	GERUND
solve	VERB	solvent	ADJECTIVE	PARTICIPLE
solve	VERB	solvent	NOUN	GERUND
te	VERB	tent	ADJECTIVE	PARTICIPLE
te	VERB	tent	NOUN	GERUND
ve	VERB	vent	ADJECTIVE	PARTICIPLE
ve	VERB	vent	NOUN	GERUND

Appendix 34

Morphological rules for "-ic" suffix

Source		Target		Relation
Morpheme	**POS**	**Morpheme**	**POS**	
a	NOUN	aic	ADJECTIVE	PERTAINER
be	NOUN	bic	ADJECTIVE	PERTAINER
bra	NOUN	braic	ADJECTIVE	PERTAINER
x	NOUN	ctic	ADJECTIVE	PERTAINER
y	NOUN	etic	ADJECTIVE	PERTAINER
fy	VERB	fic	ADJECTIVE	PARTICIPLE
a	NOUN	ic	ADJECTIVE	PERTAINER
ia	NOUN	ic	ADJECTIVE	PERTAINER
e	NOUN	ic	ADJECTIVE	PERTAINER
is	NOUN	ic	ADJECTIVE	PERTAINER
mat	NOUN	matic	ADJECTIVE	PERTAINER
m	NOUN	mmatic	ADJECTIVE	PERTAINER
n	NOUN	nic	ADJECTIVE	PERTAINER
ne	NOUN	nic	ADJECTIVE	PERTAINER
sound	NOUN	sonic	ADJECTIVE	PERTAINER
se	NOUN	stic	ADJECTIVE	PERTAINER
sis	NOUN	tic	ADJECTIVE	PERTAINER

Appendix 35

Morphological rules for "-itis" suffix

Source		Target		Relation
Morpheme	**POS**	**Morpheme**	**POS**	
x	NOUN	citis	NOUN	DISEASE
ea	NOUN	itis	NOUN	DISEASE
a	NOUN	itis	NOUN	DISEASE
y	NOUN	itis	NOUN	DISEASE
us	NOUN	itis	NOUN	DISEASE
nx	NOUN	ngitis	NOUN	DISEASE
us	NOUN	usitis	NOUN	DISEASE

Appendix 36

Complete morphological rules (final version; §5)

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
um	NOUN	a	NOUN	PLURAL	y
us	NOUN	a	NOUN	FEMININE	y
able	ADJECTIVE	ability	NOUN	ATTRIBUTE	y
ate	VERB	able	ADJECTIVE	ABLE	y
	VERB	able	ADJECTIVE	ABLE	y
ant	ADJECTIVE	able	ADJECTIVE	DERIVATIVE	y
	NOUN	able	ADJECTIVE	DERIVATIVE	n
a	NOUN	ae	NOUN	PLURAL	y
	VERB	ace	NOUN	GERUND	n
acea	NOUN	aceae	NOUN	PLURAL	n
	VERB	acy	NOUN	GERUND	n
	ADJECTIVE	ad	NOUN	QUALIFIED	n
ate	VERB	ade	NOUN	EFFECT	n
	NOUN	ade	NOUN	SUBSTANCE HOLONYM	n
	VERB	age	NOUN	GERUND	y
	NOUN	age	NOUN	DERIVATIVE	n
a	NOUN	aic	ADJECTIVE	PERTAINER	n
ain	NOUN	aincy	NOUN	GERUND OF BELIEVE PRACTICE	n
ain	VERB	aint	NOUN	GERUND	n
ate	ADJECTIVE	al	ADJECTIVE	NEARSYNONYM	y
ous	ADJECTIVE	al	ADJECTIVE	NEARSYNONYM	y
um	NOUN	al	ADJECTIVE	PERTAINER	y
on	NOUN	al	ADJECTIVE	PERTAINER	y
a	NOUN	al	ADJECTIVE	PERTAINER	n
us	NOUN	al	ADJECTIVE	PERTAINER	y
	VERB	al	NOUN	GERUND	n
al	ADJECTIVE	alise	VERB	CAUSE	y
al	ADJECTIVE	ality	NOUN	ATTRIBUTE	y
al	ADJECTIVE	alize	VERB	DERIVATIVE	y
aim	VERB	amation	NOUN	GERUND	y
	NOUN	amine	NOUN	SUBSTANCE MERONYM	n
ain	VERB	anation	NOUN	GERUND	y
a	NOUN	an	ADJECTIVE	PERTAINER	y
	NOUN	an	NOUN	INHABITANT	n
	VERB	ance	NOUN	DERIVATIVE	n
a	VERB	anda	NOUN	GERUND	n
	VERB	ando	ADJECTIVE	PARTICIPLE	n
an	ADJECTIVE	anism	NOUN	GERUND OF BELIEVE PRACTICE PERTAINYM	y
an	NOUN	anism	NOUN	GERUND OF BELIEVE PRACTICE	y
ate	VERB	ant	ADJECTIVE	PARTICIPLE	n
	VERB	ant	ADJECTIVE	PARTICIPLE	y

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
y	VERB	ant	ADJECTIVE	PARTICIPLE	n
	ADJECTIVE	ant	ADJECTIVE	NEARSYNONYM	n
ate	VERB	ant	NOUN	GERUND	n
	VERB	ant	NOUN	GERUND	n
appear	VERB	apparition	NOUN	GERUND	y
	NOUN	ar	ADJECTIVE	PERTAINER	n
	NOUN	ar	NOUN	INHABITANT	n
	NOUN	ard	NOUN	INHABITANT	n
	ADJECTIVE	ard	NOUN	QUALIFIED	n
	NOUN	ard	ADJECTIVE	QUALIFYING	n
	NOUN	ary	ADJECTIVE	ATTRIBUTE VALUE	n
	VERB	ary	ADJECTIVE	PARTICIPLE	y
a	NOUN	ary	ADJECTIVE	ATTRIBUTE VALUE	n
ate	VERB	ate	ADJECTIVE	PARTICIPLE	y
	NOUN	ate	ADJECTIVE	ATTRIBUTE VALUE	n
a	NOUN	ate	ADJECTIVE	ATTRIBUTE VALUE	n
ate	VERB	ate	NOUN	EFFECT	n
	NOUN	ate	NOUN	POSSESSION OF ATTRIBUTE	n
e	VERB	ate	VERB	NEARSYNONYM	n
a	NOUN	ate	VERB	DERIVATIVE	n
	ADJECTIVE	ate	VERB	DERIVATIVE	n
	NOUN	ate	VERB	DERIVATIVE	n
ate	VERB	ation	NOUN	GERUND	y
ise	VERB	ation	NOUN	GERUND	y
	VERB	ation	NOUN	GERUND	y
y	VERB	ation	NOUN	GERUND	y
ate	ADJECTIVE	ation	NOUN	ATTRIBUTE	y
ate	NOUN	ation	NOUN	NEARSYNONYM	y
	VERB	atious	ADJECTIVE	PARTICIPLE	y
ate	VERB	ative	ADJECTIVE	PARTICIPLE	y
	VERB	ative	ADJECTIVE	PARTICIPLE	y
ate	NOUN	ative	ADJECTIVE	PERTAINER	y
y	NOUN	ative	ADJECTIVE	PERTAINER	y
	VERB	ato	ADJECTIVE	PARTICIPLE	n
ate	VERB	ator	NOUN	SUBJECT	y
	VERB	ator	NOUN	SUBJECT	y
atory	ADJECTIVE	atory	NOUN	DERIVATIVE	y
ate	VERB	atory	ADJECTIVE	PARTICIPLE	y
	VERB	atory	ADJECTIVE	PARTICIPLE	y
b	VERB	bent	ADJECTIVE	PARTICIPLE	n
b	VERB	bent	NOUN	GERUND	n
be	NOUN	bic	ADJECTIVE	PERTAINER	n
bra	NOUN	bic	ADJECTIVE	PERTAINER	n
ble	ADJECTIVE	bilise	VERB	CAUSE	n
ble	ADJECTIVE	bly	ADVERB	PERTAINER	y
cea	NOUN	ceae	NOUN	PLURAL	n
ceive	VERB	ception	NOUN	GERUND	y
cease	VERB	cessation	NOUN	GERUND	y

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
cede	VERB	cession	NOUN	GERUND	y
ceed	VERB	cession	NOUN	GERUND	y
ce	NOUN	cial	ADJECTIVE	PERTAINER	y
cy	NOUN	cial	ADJECTIVE	PERTAINER	n
x	NOUN	cial	ADJECTIVE	PERTAINER	n
t	NOUN	cial	ADJECTIVE	PERTAINER	n
ce	VERB	cion	NOUN	GERUND	n
x	NOUN	citis	NOUN	DISEASE	n
construct	VERB	construction	NOUN	GERUND	y
construe	VERB	construction	NOUN	GERUND	y
ct	NOUN	ctal	ADJECTIVE	PERTAINER	n
x	NOUN	ctic	ADJECTIVE	PERTAINER	n
ct	VERB	ction	NOUN	GERUND	y
ct	ADJECTIVE	ction	NOUN	ATTRIBUTE	n
t	ADJECTIVE	cy	NOUN	GERUND OF BELIEVE PRACTICE PERTAINYM	n
t	NOUN	cy	NOUN	GERUND OF BELIEVE PRACTICE	n
te	ADJECTIVE	cy	NOUN	GERUND OF BELIEVE PRACTICE PERTAINYM	n
d	NOUN	dal	ADJECTIVE	PERTAINER	n
de	NOUN	dal	ADJECTIVE	PERTAINER	n
	NOUN	de	ADJECTIVE	PERTAINER	n
	NOUN	de	NOUN	SUBSTANCE MERONYM	n
define	VERB	definition	NOUN	GERUND	y
deem	VERB	demption	NOUN	GERUND	y
de	VERB	dent	ADJECTIVE	PARTICIPLE	n
de	VERB	dent	NOUN	GERUND	n
dge	VERB	dgment	NOUN	GERUND	y
de	NOUN	dinal	ADJECTIVE	PERTAINER	n
	NOUN	dom	NOUN	POSSESSION OF ATTRIBUTE	y
duke	NOUN	ducal	ADJECTIVE	PERTAINER	n
ea	NOUN	eae	NOUN	PLURAL	y
ea	NOUN	eal	ADJECTIVE	PERTAINER	n
e	NOUN	ear	ADJECTIVE	PERTAINER	n
	NOUN	ed	ADJECTIVE	ATTRIBUTE VALUE	y
	VERB	ee	NOUN	PATIENT	n
	NOUN	eer	NOUN	SUBJECT OF BELIEVE PRACTICE	n
	NOUN	el	NOUN	DIMINUTIVE	n
	NOUN	ella	NOUN	DIMINUTIVE	n
e	NOUN	ely	ADJECTIVE	ATTRIBUTE VALUE	y
	ADJECTIVE	en	VERB	DERIVATIVE	y
	NOUN	en	VERB	CAUSE	n
	NOUN	en	ADJECTIVE	PERTAINER	n
ent	NOUN	entary	ADJECTIVE	PERTAINER	y
e	ADJECTIVE	eous	ADJECTIVE	NEARSYNONYM	y

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
y	NOUN	eous	ADJECTIVE	PERTAINER	n
	VERB	er	NOUN	SUBJECT	y
	NOUN	er	NOUN	INHABITANT	n
	VERB	er	VERB	NEARSYNONYM	y
er	VERB	erent	ADJECTIVE	PARTICIPLE	n
ere	VERB	erent	ADJECTIVE	PARTICIPLE	n
er	VERB	erent	NOUN	GERUND	n
ere	VERB	erent	NOUN	GERUND	n
	VERB	ery	NOUN	DERIVATIVE	n
	NOUN	ery	NOUN	DERIVATIVE	n
er	NOUN	ery	NOUN	DERIVATIVE	n
er	VERB	ery	NOUN	DERIVATIVE	n
	NOUN	esque	ADJECTIVE	RESEMBLING	n
	ADJECTIVE	esque	ADJECTIVE	NEARSYNONYM	n
	NOUN	ess	NOUN	FEMININE	n
ess	VERB	essant	ADJECTIVE	PARTICIPLE	n
eed	VERB	essive	NOUN	GERUND	n
	NOUN	et	NOUN	DIMINUTIVE	n
y	NOUN	etic	ADJECTIVE	PERTAINER	n
ete	VERB	etion	NOUN	GERUND	y
ete	ADJECTIVE	etion	NOUN	ATTRIBUTE	n
	NOUN	ette	NOUN	DIMINUTIVE	n
e	ADJECTIVE	ety	NOUN	ATTRIBUTE	y
ect	VERB	exion	NOUN	GERUND	y
fy	VERB	faction	NOUN	GERUND	y
fy	VERB	fic	ADJECTIVE	PARTICIPLE	n
fy	VERB	fication	NOUN	GERUND	y
	NOUN	form	ADJECTIVE	RESEMBLING	n
form	ADJECTIVE	form	NOUN	ATTRIBUTE	n
	NOUN	ful	NOUN	MEASUREDBY	y
	NOUN	ful	ADJECTIVE	ATTRIBUTE VALUE	y
	VERB	ful	ADJECTIVE	PARTICIPLE	y
ge	VERB	gent	ADJECTIVE	PARTICIPLE	n
ge	VERB	gent	NOUN	GERUND	n
h	NOUN	hal	ADJECTIVE	PERTAINER	n
here	VERB	hesion	NOUN	GERUND	y
her	VERB	hesion	NOUN	GERUND	y
	NOUN	hood	NOUN	POSSESSION OF ATTRIBUTE	y
	ADJECTIVE	hood	NOUN	ATTRIBUTE	n
us	NOUN	i	NOUN	PLURAL	y
ium	NOUN	ia	NOUN	PLURAL	y
iacea	NOUN	iaceae	NOUN	PLURAL	n
	NOUN	ial	ADJECTIVE	PERTAINER	n
us	NOUN	ian	ADJECTIVE	PERTAINER	n
y	NOUN	ian	NOUN	SUBJECT OF BELIEVE PRACTICE	n
	NOUN	ian	NOUN	SUBJECT OF BELIEVE PRACTICE	y
	ADJECTIVE	ian	NOUN	SUBJECT OF BELIEVE	y

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
				PRACTICE PERTAINYM	
y	VERB	iant	ADJECTIVE	PARTICIPLE	y
y	VERB	iant	NOUN	GERUND	y
ible	ADJECTIVE	ibility	NOUN	ATTRIBUTE	y
	VERB	ible	ADJECTIVE	ABLE	y
ion	NOUN	ible	ADJECTIVE	ABILITY	n
	NOUN	ic	NOUN	DERIVATIVE	n
y	NOUN	ic	ADJECTIVE	PERTAINER	n
ise	VERB	ic	ADJECTIVE	DERIVATIVE	y
ize	VERB	ic	ADJECTIVE	DERIVATIVE	y
a	NOUN	ic	ADJECTIVE	PERTAINER	n
ia	NOUN	ic	ADJECTIVE	PERTAINER	n
e	NOUN	ic	ADJECTIVE	PERTAINER	n
is	NOUN	ic	ADJECTIVE	PERTAINER	n
ic	ADJECTIVE	ical	ADJECTIVE	SYNONYM	y
ic	NOUN	ical	ADJECTIVE	PERTAINER	y
ics	NOUN	ical	ADJECTIVE	PERTAINER	y
y	NOUN	ical	ADJECTIVE	DERIVATIVE	y
ex	NOUN	ical	ADJECTIVE	DERIVATIVE	y
ix	NOUN	ical	ADJECTIVE	DERIVATIVE	y
	NOUN	ical	ADJECTIVE	PERTAINER	n
y	NOUN	ical	ADJECTIVE	PERTAINER	n
ice	NOUN	ical	ADJECTIVE	PERTAINER	n
ical	ADJECTIVE	ical	NOUN	QUALIFIED	y
ical	ADJECTIVE	ically	ADVERB	PERTAINER	y
ic	ADJECTIVE	ically	ADVERB	PERTAINER	y
y	VERB	ication	NOUN	GERUND	y
y	VERB	icator	NOUN	SUBJECT	y
ise	VERB	ice	NOUN	GERUND	n
	NOUN	ice	NOUN	GERUND OF BELIEVE PRACTICE	n
y	NOUN	ician	NOUN	SUBJECT OF BELIEVE PRACTICE	y
ic	ADJECTIVE	ician	NOUN	SUBJECT OF BELIEVE PRACTICE PERTAINYM	y
ic	NOUN	ician	NOUN	SUBJECT OF BELIEVE PRACTICE	y
ics	NOUN	ician	NOUN	SUBJECT OF BELIEVE PRACTICE	y
ics	NOUN	icist	NOUN	SUBJECT OF BELIEVE PRACTICE	y
	NOUN	icle	NOUN	DIMINUTIVE	n
ic	ADJECTIVE	ics	NOUN	QUALIFIED	n
	NOUN	id	ADJECTIVE	QUALIFYING	n
	ADJECTIVE	id	NOUN	QUALIFIED	y
id	NOUN	ida	NOUN	FEMININE	n
ida	NOUN	idae	NOUN	PLURAL	n
idise	VERB	idant	ADJECTIVE	PARTICIPLE	n

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
idise	VERB	idant	NOUN	GERUND	n
	NOUN	ide	ADJECTIVE	PERTAINER	n
	NOUN	ide	NOUN	SUBSTANCE MERONYM	n
id	ADJECTIVE	idea	NOUN	PERTAINYM	n
y	NOUN	ie	NOUN	SYNONYM	y
ier	NOUN	iere	NOUN	FEMININE	n
	NOUN	iferous	ADJECTIVE	QUALIFYING	n
	NOUN	iform	ADJECTIVE	RESEMBLING	n
iform	ADJECTIVE	iform	NOUN	ATTRIBUTE	n
iform	NOUN	iformes	NOUN	PLURAL	n
	ADJECTIVE	ify	VERB	DERIVATIVE	n
e	ADJECTIVE	ify	VERB	DERIVATIVE	n
	NOUN	ify	VERB	DERIVATIVE	n
e	NOUN	ify	VERB	DERIVATIVE	n
	NOUN	il	NOUN	DIMINUTIVE	n
	NOUN	il	ADJECTIVE	QUALIFYING	n
	NOUN	illa	NOUN	DIMINUTIVE	n
ile	ADJECTIVE	ility	NOUN	ATTRIBUTE	y
	NOUN	in	NOUN	SUBSTANCE MERONYM	n
	ADJECTIVE	in	NOUN	ATTRIBUTE	n
ina	NOUN	inae	NOUN	PLURAL	n
	NOUN	inant	NOUN	DIMINUTIVE	n
in	VERB	inant	ADJECTIVE	PARTICIPLE	n
in	VERB	inant	NOUN	GERUND	n
	NOUN	ine	ADJECTIVE	PERTAINER	n
	NOUN	ine	NOUN	SUBSTANCE MERONYM	n
	ADJECTIVE	ine	NOUN	ATTRIBUTE	n
ain	VERB	inent	ADJECTIVE	PARTICIPLE	n
ain	VERB	inent	NOUN	GERUND	n
on	NOUN	ino	NOUN	DIMINUTIVE	n
ion	NOUN	ional	ADJECTIVE	PERTAINER	y
ion	NOUN	ionary	ADJECTIVE	PERTAINER	y
ion	NOUN	ionary	NOUN	SUBJECT OF VERBSOURCE OF GERUND	y
y	NOUN	ious	ADJECTIVE	PERTAINER	n
ise	VERB	isation	NOUN	GERUND	y
	NOUN	is	NOUN	DERIVATIVE	n
	ADJECTIVE	ise	VERB	CAUSE	n
	NOUN	ise	VERB	CAUSE	n
y	NOUN	ise	VERB	BELIEVE PRACTICE	y
	NOUN	ish	ADJECTIVE	PERTAINER	y
	ADJECTIVE	ish	ADJECTIVE	DIMINUTIVE	y
ise	ADJECTIVE	ision	NOUN	ATTRIBUTE	y
ise	VERB	ism	NOUN	GERUND	y
	NOUN	ism	NOUN	GERUND OF BELIEVE PRACTICE	y
	ADJECTIVE	ism	NOUN	GERUND OF BELIEVE PRACTICE	y

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
				PERTAINYM	
	VERB	ism	NOUN	GERUND	n
ist	NOUN	ist	ADJECTIVE	PERTAINER	n
y	NOUN	ist	NOUN	SUBJECT OF BELIEVE PRACTICE	n
	ADJECTIVE	ist	NOUN	SUBJECT OF BELIEVE PRACTICE PERTAINYM	n
	NOUN	ist	NOUN	SUBJECT OF BELIEVE PRACTICE	n
	VERB	ist	NOUN	SUBJECT	n
a	NOUN	ist	NOUN	SUBJECT OF BELIEVE PRACTICE	n
ism	NOUN	ist	NOUN	SUBJECT OF VERBSOURCE OF GERUND	y
ist	VERB	istent	ADJECTIVE	PARTICIPLE	n
ist	VERB	istent	NOUN	GERUND	n
ist	NOUN	istic	ADJECTIVE	PERTAINER	n
it	NOUN	ital	ADJECTIVE	PERTAINER	n
e	VERB	ite	ADJECTIVE	DERIVATIVE	n
	NOUN	ite	NOUN	INHABITANT	n
ise	VERB	ition	NOUN	GERUND	y
ize	VERB	ition	NOUN	GERUND	y
ish	VERB	ition	NOUN	GERUND	y
ite	ADJECTIVE	ition	NOUN	ATTRIBUTE	y
ea	NOUN	itis	NOUN	DISEASE	n
a	NOUN	itis	NOUN	DISEASE	n
y	NOUN	itis	NOUN	DISEASE	n
us	NOUN	itis	NOUN	DISEASE	n
itt	VERB	ittent	ADJECTIVE	PARTICIPLE	n
itt	VERB	ittent	NOUN	GERUND	n
	ADJECTIVE	itude	NOUN	ATTRIBUTE	y
ous	ADJECTIVE	ity	NOUN	ATTRIBUTE	y
ious	ADJECTIVE	ity	NOUN	ATTRIBUTE	y
e	ADJECTIVE	ity	NOUN	ATTRIBUTE	y
	ADJECTIVE	ity	NOUN	ATTRIBUTE	y
al	ADJECTIVE	ity	NOUN	ATTRIBUTE	y
	VERB	ity	NOUN	GERUND	n
	NOUN	ium	NOUN	SUBSTANCE MERONYM	n
	ADJECTIVE	ium	NOUN	ATTRIBUTE	n
ive	NOUN	ival	ADJECTIVE	PERTAINER	n
ive	ADJECTIVE	ival	ADJECTIVE	NEARSYNONYM	n
	VERB	ive	ADJECTIVE	PARTICIPLE	n
ion	NOUN	ive	ADJECTIVE	PERTAINER	n
ive	ADJECTIVE	ive	NOUN	QUALIFIED	y
ize	VERB	ization	NOUN	DERIVATIVE	y
	NOUN	ize	VERB	DERIVATIVE	y
y	NOUN	ize	VERB	BELIEVE PRACTICE	y
ise	VERB	ize	VERB	SYNONYM	y

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
join	VERB	junction	NOUN	GERUND	y
know	VERB	knowledge	NOUN	GERUND	y
ll	VERB	lant	ADJECTIVE	PARTICIPLE	n
ll	VERB	lant	NOUN	GERUND	n
ll	VERB	lent	ADJECTIVE	PARTICIPLE	n
	NOUN	le	NOUN	DIMINUTIVE	n
ll	VERB	lent	NOUN	GERUND	n
	NOUN	less	ADJECTIVE	ANTONYM OF ATTRIBUTE VALUE	y
	VERB	less	ADJECTIVE	ANTONYM OF PARTICIPLE	y
	NOUN	let	NOUN	DIMINUTIVE	n
	NOUN	like	ADJECTIVE	PERTAINER	y
	NOUN	ling	NOUN	DIMINUTIVE	n
le	ADJECTIVE	lity	NOUN	QUALIFIED	y
l	VERB	ll	VERB	SYNONYM	y
l	VERB	llent	ADJECTIVE	PARTICIPLE	n
l	VERB	llent	NOUN	GERUND	n
l	VERB	llion	NOUN	GERUND	n
le	NOUN	ly	ADJECTIVE	ATTRIBUTE VALUE	y
	NOUN	ly	ADJECTIVE	ATTRIBUTE VALUE	n
l	NOUN	ly	ADJECTIVE	ATTRIBUTE VALUE	n
	ADJECTIVE	ly	ADJECTIVE	NEARSYNONYM	y
	ADJECTIVE	ly	ADVERB	PERTAINER	y
le	VERB	ly	NOUN	GERUND	n
m	NOUN	mal	ADJECTIVE	PERTAINER	n
mat	NOUN	matic	ADJECTIVE	PERTAINER	y
ma	NOUN	matic	ADJECTIVE	PERTAINER	y
m	NOUN	matic	ADJECTIVE	PERTAINER	y
ma	NOUN	matise	VERB	CAUSE	n
	VERB	ment	NOUN	DERIVATIVE	y
er	VERB	ment	NOUN	DERIVATIVE	y
merge	VERB	mersion	NOUN	GERUND	n
mit	VERB	mission	NOUN	GERUND	y
m	NOUN	mmatic	ADJECTIVE	PERTAINER	n
move	VERB	motion	NOUN	GERUND	y
n	NOUN	na	NOUN	FEMININE	n
num	NOUN	na	NOUN	PLURAL	n
ne	NOUN	nal	ADJECTIVE	PERTAINER	n
n	NOUN	nal	ADJECTIVE	PERTAINER	n
nt	ADJECTIVE	nce	NOUN	ATTRIBUTE	y
nt	VERB	nce	NOUN	GERUND	n
nce	NOUN	ncial	ADJECTIVE	PERTAINER	y
nt	ADJECTIVE	ncy	NOUN	ATTRIBUTE	y
nd	VERB	ndant	ADJECTIVE	PARTICIPLE	n
nd	VERB	ndant	NOUN	GERUND	n
nd	VERB	ndent	ADJECTIVE	PARTICIPLE	n
nd	VERB	ndent	NOUN	GERUND	n
	NOUN	ne	NOUN	SUBSTANCE MERONYM	n

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
	ADJECTIVE	ne	NOUN	ATTRIBUTE	n
neglect	VERB	negligent	ADJECTIVE	PARTICIPLE	n
	ADJECTIVE	ness	NOUN	ATTRIBUTE	y
nx	NOUN	ngeal	ADJECTIVE	PERTAINER	n
nx	NOUN	ngitis	NOUN	DISEASE	n
n	NOUN	nic	ADJECTIVE	PERTAINER	n
ne	NOUN	nic	ADJECTIVE	PERTAINER	n
nd	VERB	nsion	NOUN	GERUND	n
nd	VERB	nsive	ADJECTIVE	PARTICIPLE	n
nt	ADJECTIVE	nt	NOUN	QUALIFIED	y
nt	NOUN	ntal	ADJECTIVE	PERTAINER	n
nce	NOUN	ntial	ADJECTIVE	PERTAINER	y
nt	NOUN	ntial	ADJECTIVE	PERTAINER	y
nce	NOUN	ntial	NOUN	DERIVATIVE	y
nce	NOUN	ntiate	VERB	DEMONSTRATE	y
nt	ADJECTIVE	ntiate	VERB	DERIVATIVE	y
nd	VERB	ntion	NOUN	GERUND	y
nounce	VERB	nunciation	NOUN	GERUND	y
nourish	VERB	nutrition	NOUN	GERUND	y
	NOUN	o	NOUN	DERIVATIVE	n
obey	VERB	obedient	ADJECTIVE	PARTICIPLE	n
oke	VERB	ocation	NOUN	GERUND	y
	NOUN	oid	ADJECTIVE	RESEMBLING	y
	NOUN	oid	NOUN	RESEMBLING	y
oid	ADJECTIVE	oidea	NOUN	PERTAINYM	n
	NOUN	ol	NOUN	SUBSTANCE MERONYM	n
	NOUN	ology	NOUN	GERUND OF BELIEVE PRACTICE	n
a	NOUN	ology	NOUN	GERUND OF BELIEVE PRACTICE	n
olve	VERB	olution	NOUN	GERUND	y
	NOUN	on	NOUN	SUBSTANCE MERONYM	n
	ADJECTIVE	on	NOUN	ATTRIBUTE	n
	NOUN	one	NOUN	SUBSTANCE MERONYM	n
	ADJECTIVE	one	NOUN	ATTRIBUTE	n
ound	VERB	onent	ADJECTIVE	PARTICIPLE	n
ose	VERB	onent	ADJECTIVE	PARTICIPLE	n
ound	VERB	onent	NOUN	GERUND	n
ose	VERB	onent	NOUN	GERUND	n
onium	NOUN	onia	NOUN	PLURAL	n
on	NOUN	onia	NOUN	POSSESSION OF ATTRIBUTE	n
onic	ADJECTIVE	onia	NOUN	ATTRIBUTE	n
	VERB	or	NOUN	SUBJECT	y
or	NOUN	orate	NOUN	POSSESSION OF ATTRIBUTE	y
or	NOUN	orial	ADJECTIVE	PERTAINER	y
orb	VERB	orption	NOUN	GERUND	y
ion	NOUN	ory	ADJECTIVE	PERTAINER	y
	VERB	ory	ADJECTIVE	PARTICIPLE	n

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
	NOUN	ose	ADJECTIVE	PERTAINER	n
	NOUN	ose	NOUN	SUBSTANCE MERONYM	n
ose	VERB	osition	NOUN	GERUND	y
ous	ADJECTIVE	osity	NOUN	ATTRIBUTE	y
	NOUN	ous	ADJECTIVE	PERTAINER	n
e	VERB	ous	ADJECTIVE	PARTICIPLE	n
	VERB	ous	ADJECTIVE	PARTICIPLE	n
y	NOUN	ous	ADJECTIVE	PERTAINER	n
on	NOUN	ous	ADJECTIVE	PERTAINER	n
ic	ADJECTIVE	ous	ADJECTIVE	NEARSYNONYM	n
ove	VERB	oval	NOUN	GERUND	n
pe	NOUN	pal	ADJECTIVE	PERTAINER	n
peat	VERB	petition	NOUN	GERUND	y
pete	VERB	petition	NOUN	GERUND	y
pose	VERB	posal	NOUN	GERUND	n
prove	VERB	probation	NOUN	GERUND	y
pel	VERB	pulsion	NOUN	GERUND	y
quire	VERB	quisition	NOUN	GERUND	y
re	NOUN	ral	ADJECTIVE	PERTAINER	n
er	NOUN	ral	ADJECTIVE	PERTAINER	n
ra	NOUN	ral	ADJECTIVE	PERTAINER	n
or	NOUN	ral	ADJECTIVE	PERTAINER	n
r	NOUN	ral	ADJECTIVE	PERTAINER	n
er	VERB	rance	NOUN	GERUND	y
er	VERB	rant	ADJECTIVE	PARTICIPLE	n
re	VERB	rant	ADJECTIVE	PARTICIPLE	n
er	VERB	rant	NOUN	GERUND	n
re	VERB	rant	NOUN	GERUND	n
rd	VERB	rdant	ADJECTIVE	PARTICIPLE	n
rd	VERB	rdant	NOUN	GERUND	n
render	VERB	rendition	NOUN	GERUND	y
rete	VERB	retion	NOUN	GERUND	y
r	NOUN	rial	ADJECTIVE	PERTAINER	n
rr	VERB	rrent	ADJECTIVE	PARTICIPLE	n
r	VERB	rrent	ADJECTIVE	PARTICIPLE	n
rr	VERB	rrent	NOUN	GERUND	n
r	VERB	rrent	NOUN	GERUND	n
rt	VERB	rsion	NOUN	GERUND	y
rt	ADJECTIVE	rsion	NOUN	ATTRIBUTE	n
er	VERB	ry	NOUN	GERUND	y
	NOUN	ry	NOUN	DERIVATIVE	y
	NOUN	s	NOUN	PLURAL	y
se	NOUN	sal	ADJECTIVE	PERTAINER	n
ss	NOUN	sal	ADJECTIVE	PERTAINER	n
save	VERB	salvation	NOUN	GERUND	y
se	VERB	sant	ADJECTIVE	PARTICIPLE	n
se	VERB	sant	NOUN	GERUND	n
sce	VERB	scent	ADJECTIVE	PARTICIPLE	n
sce	VERB	scent	NOUN	GERUND	n
scribe	VERB	scription	NOUN	GERUND	y

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
sense	VERB	sentient	ADJECTIVE	PARTICIPLE	n
sense	VERB	sentient	NOUN	GERUND	n
sent	VERB	sension	NOUN	GERUND	y
sense	VERB	sensitive	ADJECTIVE	PARTICIPLE	n
	NOUN	ship	NOUN	POSSESSION OF ATTRIBUTE	y
	ADJECTIVE	ship	NOUN	ATTRIBUTE	y
d	VERB	sible	ADJECTIVE	DERIVATIVE	n
se	VERB	sion	NOUN	GERUND	y
de	VERB	sion	NOUN	GERUND	y
solve	VERB	solvent	ADJECTIVE	PARTICIPLE	n
solve	VERB	solvent	NOUN	GERUND	n
	NOUN	some	ADJECTIVE	PERTAINER	y
	VERB	some	ADJECTIVE	PARTICIPLE	y
	ADJECTIVE	some	ADJECTIVE	NEARSYNONYM	y
sound	NOUN	sonic	ADJECTIVE	PERTAINER	n
spoil	VERB	spoliation	NOUN	GERUND	y
	NOUN	sque	ADJECTIVE	RESEMBLING	n
	ADJECTIVE	sque	ADJECTIVE	NEARSYNONYM	n
ss	VERB	ssion	NOUN	GERUND	y
st	NOUN	stal	ADJECTIVE	PERTAINER	n
se	NOUN	stic	ADJECTIVE	PERTAINER	n
suck	VERB	suction	NOUN	GERUND	y
suspect	VERB	suspicion	NOUN	GERUND	y
t	VERB	tant	ADJECTIVE	PARTICIPLE	n
te	VERB	tant	ADJECTIVE	PARTICIPLE	n
t	VERB	tant	NOUN	GERUND	n
te	VERB	tant	NOUN	GERUND	n
ty	NOUN	tarian	NOUN	SUBJECT OF VERBSOURCE OF GERUND	y
te	VERB	tent	ADJECTIVE	PARTICIPLE	n
te	VERB	tent	NOUN	GERUND	n
ty	NOUN	tal	ADJECTIVE	PERTAINER	n
	VERB	te	ADJECTIVE	DERIVATIVE	n
	ADJECTIVE	th	ADJECTIVE	REPETITION	y
ce	NOUN	tial	ADJECTIVE	PERTAINER	y
cy	NOUN	tial	ADJECTIVE	PERTAINER	n
t	NOUN	tial	ADJECTIVE	PERTAINER	n
sis	NOUN	tic	ADJECTIVE	PERTAINER	y
te	VERB	tion	NOUN	GERUND	y
t	VERB	tion	NOUN	GERUND	y
ce	NOUN	tist	NOUN	SUBJECT OF BELIEVE PRACTICE	n
ce	ADJECTIVE	tive	ADJECTIVE	DERIVATIVE	y
te	ADJECTIVE	tive	ADJECTIVE	DERIVATIVE	y
t	VERB	tor	NOUN	DERIVATIVE	y
t	VERB	ttal	NOUN	GERUND	n
t	VERB	ture	NOUN	GERUND	n
	ADJECTIVE	ty	NOUN	ATTRIBUTE	n
	NOUN	ual	ADJECTIVE	PERTAINER	n

Source		Target		Relation	Applicable to monosyllables?
Morpheme	**POS**	**Morpheme**	**POS**		
ue	VERB	uant	ADJECTIVE	PARTICIPLE	n
ue	VERB	uant	NOUN	GERUND	n
	NOUN	uate	VERB	DERIVATIVE	n
uce	VERB	uction	NOUN	GERUND	y
ude	NOUN	udinal	ADJECTIVE	PERTAINER	n
	NOUN	ula	NOUN	DIMINUTIVE	n
le	NOUN	ular	ADJECTIVE	PERTAINER	y
le	NOUN	ulate	ADJECTIVE	ATTRIBUTE VALUE	y
le	NOUN	ulate	VERB	CAUSE	y
le	NOUN	ulous	ADJECTIVE	PERTAINER	n
	NOUN	um	NOUN	SUBSTANCE MERONYM	n
	ADJECTIVE	um	NOUN	ATTRIBUTE	n
ume	VERB	umption	NOUN	GERUND	y
ounce	VERB	unciant	ADJECTIVE	PARTICIPLE	n
ounce	VERB	unciant	NOUN	GERUND	n
ur	VERB	ursion	NOUN	GERUND	y
ound	VERB	undant	NOUN	GERUND	n
	VERB	ure	NOUN	GERUND	n
	VERB	urus	NOUN	GERUND	n
	NOUN	us	NOUN	DERIVATIVE	n
us	NOUN	usitis	NOUN	DISEASE	n
ude	VERB	usive	ADJECTIVE	PARTICIPLE	n
ute	ADJECTIVE	ution	NOUN	ATTRIBUTE	y
ve	VERB	vant	ADJECTIVE	PARTICIPLE	n
ve	VERB	vant	NOUN	GERUND	n
ve	VERB	vent	ADJECTIVE	PARTICIPLE	n
ve	VERB	vent	NOUN	GERUND	n
vene	VERB	vention	NOUN	GERUND	y
verb	NOUN	verbial	ADJECTIVE	PERTAINER	n
w	VERB	wal	NOUN	GERUND	n
	NOUN	ward	ADVERB	DIRECTION	n
ward	ADVERB	wards	ADVERB	SYNONYM	y
	ADJECTIVE	ware	NOUN	QUALIFIED	y
	NOUN	ware	NOUN	SUBSTANCE HOLONYM	y
	VERB	ware	NOUN	SUBJECT	y
	ADJECTIVE	wise	ADVERB	PERTAINER	y
	NOUN	wise	ADVERB	PERTAINER	y
c	NOUN	x	NOUN	DERIVATIVE	n
g	NOUN	x	NOUN	DERIVATIVE	n
	NOUN	y	ADJECTIVE	ATTRIBUTE VALUE	n
e	NOUN	y	ADJECTIVE	DERIVATIVE	y
	VERB	y	ADJECTIVE	DERIVATIVE	n
	ADJECTIVE	y	NOUN	DERIVATIVE	n
	NOUN	yl	ADJECTIVE	PERTAINER	n
yse	VERB	ysate	NOUN	EFFECT	y
yse	VERB	ysis	NOUN	GERUND	y
yse	VERB	yze	VERB	SYNONYM	y
	ADJECTIVE		ADVERB	PERTAINER	y

Source		Target		Relation	Applicable to monosyllables?
Morpheme	POS	Morpheme	POS		
	ADVERB		ADJECTIVE	PERTAINYM	y
	ADJECTIVE		NOUN	DERIV	n
	VERB		NOUN	DERIV	n
	NOUN		VERB	DERIV	n
	NOUN		ADJECTIVE	DERIV	n
	PREPOSITION		ADVERB	DERIV	y
	ADVERB		PREPOSITION	DERIV	y

Appendix 37

Primary suffixation analysis results for "-able", "-ical" & "-ician"

Original word	Original POS	Desuffixed word	Desuffixed POS	Relation type
academician	NOUN	academic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
acoustician	NOUN	acoustic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
aesthetician	NOUN	aesthetic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
cosmetician	NOUN	cosmetic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
diagnostician	NOUN	diagnostic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
econometrician	NOUN	econometric	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
electrician	NOUN	electric	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
esthetician	NOUN	esthetic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
geometrician	NOUN	geometric	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
geriatrician	NOUN	geriatric	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
logistician	NOUN	logistic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
obstetrician	NOUN	obstetric	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
optician	NOUN	optic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
paediatrician	NOUN	paediatric	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
pediatrician	NOUN	pediatric	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
phonetician	NOUN	phonetic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
semiotician	NOUN	semiotic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
syntactician	NOUN	syntactic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
theoretician	NOUN	theoretic	ADJECTIVE	PERTAINER TO OBJECT OF BELIEF PRACTICE OF ROLE
arithmetician	NOUN	arithmetic	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
clinician	NOUN	clinic	NOUN	OBJECT OF BELIEF

Original word	Original POS	Desuffixed word	Desuffixed POS	Relation type
				PRACTICE OF ROLE
dialectician	NOUN	dialectic	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
ethician	NOUN	ethic	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
logician	NOUN	logic	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
magician	NOUN	magic	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
musician	NOUN	music	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
rhetorician	NOUN	rhetoric	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
statistician	NOUN	statistic	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
tactician	NOUN	tactic	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
mathematician	NOUN	mathematics	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
physician	NOUN	physics	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
politician	NOUN	politics	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
beautician	NOUN	beauty	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
photometrician	NOUN	photometry	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
trigonometrician	NOUN	trigonometry	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
dietician	NOUN	diet	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE
*patrician	NOUN	pater	NOUN	OBJECT OF BELIEF PRACTICE OF ROLE

Appendix 38

False lexical stems (Prefixation stem stoplist)

Stem	POS	Stem	POS	Stem	POS
a	NOUN	cardia	NOUN	den	NOUN
ace	NOUN	carp	NOUN	dent	NOUN
ad	NOUN	carpus	NOUN	dent	VERB
ade	NOUN	caustic	NOUN	denture	NOUN
age	VERB	cay	NOUN	derma	NOUN
age	NOUN	cede	VERB	don	NOUN
agio	NOUN	cent	NOUN	don	VERB
aldol	NOUN	cert	NOUN	dopa	NOUN
amide	NOUN	chase	NOUN	drawn	ADJECTIVE
amine	NOUN	chase	VERB	dress	NOUN
amnios	NOUN	cheat	NOUN	dress	VERB
angel	NOUN	chequer	NOUN	drome	NOUN
ant	NOUN	chief	NOUN	duce	NOUN
apse	NOUN	china	NOUN	duct	NOUN
apsis	NOUN	chore	NOUN	dural	ADJECTIVE
ar	NOUN	chorea	NOUN	e	NOUN
arch	NOUN	chrome	NOUN	el	NOUN
as	ADVERB	chrome	VERB	en	NOUN
assay	NOUN	cilium	NOUN	ern	NOUN
assay	VERB	cite	VERB	ex	ADJECTIVE
aster	NOUN	claim	NOUN	fair	NOUN
at	NOUN	claim	VERB	feat	NOUN
avo	NOUN	clast	NOUN	fence	NOUN
ax	NOUN	clonal	ADJECTIVE	fice	NOUN
ax	VERB	clonus	NOUN	file	NOUN
bat	NOUN	cocci	NOUN	file	VERB
bat	VERB	coccus	NOUN	fine	NOUN
bate	VERB	col	NOUN	fine	VERB
bet	VERB	comb	NOUN	fine	ADJECTIVE
bettor	NOUN	come	NOUN	firm	VERB
biotic	ADJECTIVE	company	VERB	fit	NOUN
blast	NOUN	compass	VERB	fit	VERB
bola	NOUN	con	NOUN	flavin	NOUN
bole	NOUN	cope	NOUN	flex	NOUN
boss	VERB	cord	NOUN	flex	VERB
brace	NOUN	cord	VERB	flux	NOUN
brace	VERB	corn	NOUN	ford	VERB
bridge	VERB	cost	VERB	form	NOUN
broider	VERB	cot	NOUN	form	VERB
buff	NOUN	cote	NOUN	fort	NOUN
buff	VERB	counter	NOUN	found	VERB
bunk	VERB	counter	VERB	found	ADJECTIVE
bust	VERB	crescent	ADJECTIVE	fray	NOUN
cadent	ADJECTIVE	critic	NOUN	fray	VERB
cant	VERB	cullis	NOUN	fringe	VERB
canthus	NOUN	cumber	VERB	fuddle	VERB
cape	NOUN	cure	VERB	fugal	ADJECTIVE
card	NOUN	cuss	VERB	furan	NOUN
card	VERB	d	ADJECTIVE	fuse	VERB

Stem	POS	Stem	POS	Stem	POS
fusion	NOUN	lexis	NOUN	on	ADJECTIVE
gam	NOUN	li	NOUN	one	NOUN
gauss	NOUN	liberate	VERB	opsin	NOUN
gavage	NOUN	ligate	VERB	os	NOUN
gee	NOUN	light	NOUN	over	NOUN
gee	VERB	light	VERB	overt	ADJECTIVE
gen	NOUN	light	ADJECTIVE	pact	NOUN
genic	ADJECTIVE	lime	VERB	pal	VERB
genital	ADJECTIVE	lite	ADJECTIVE	pale	VERB
glut	VERB	literate	ADJECTIVE	pall	VERB
gnosis	NOUN	log	NOUN	pane	NOUN
gnostic	ADJECTIVE	long	VERB	pare	VERB
gram	NOUN	lope	VERB	pat	NOUN
gramme	NOUN	lucent	ADJECTIVE	pause	NOUN
gross	VERB	luge	NOUN	pe	NOUN
gust	NOUN	luge	VERB	peach	VERB
habit	VERB	lysin	NOUN	peal	NOUN
hale	VERB	lysis	NOUN	peal	VERB
hap	NOUN	m	NOUN	pediment	NOUN
hash	VERB	ma	NOUN	pert	ADJECTIVE
hectic	ADJECTIVE	mantic	ADJECTIVE	pet	NOUN
hemin	NOUN	mantle	VERB	petite	NOUN
hen	NOUN	mark	NOUN	phage	NOUN
hod	NOUN	mat	NOUN	philia	NOUN
hyalin	NOUN	mate	NOUN	phone	NOUN
ic	ADJECTIVE	mate	VERB	phony	NOUN
icky	ADJECTIVE	mend	VERB	pia	NOUN
id	NOUN	mere	NOUN	pile	VERB
in	NOUN	metric	ADJECTIVE	pilous	ADJECTIVE
in	ADVERB	mezzo	NOUN	plain	VERB
in	PREPOSITION	mire	VERB	plant	VERB
ion	NOUN	miss	VERB	plasm	NOUN
iritis	NOUN	mite	NOUN	plate	NOUN
ism	NOUN	mo	NOUN	plica	NOUN
jig	VERB	mode	NOUN	ploy	NOUN
juror	NOUN	mons	NOUN	ply	VERB
jury	NOUN	moron	NOUN	ply	NOUN
kinase	NOUN	mum	NOUN	pod	NOUN
kine	NOUN	mum	ADJECTIVE	podium	NOUN
kinin	NOUN	mural	ADJECTIVE	point	NOUN
l	NOUN	mute	VERB	point	VERB
l	ADJECTIVE	mute	NOUN	port	NOUN
la	NOUN	n	NOUN	port	VERB
labile	ADJECTIVE	native	NOUN	pose	NOUN
lapidate	VERB	native	ADJECTIVE	pose	VERB
lapse	NOUN	nine	NOUN	posit	NOUN
lapse	VERB	novate	VERB	posit	VERB
lard	VERB	nuncio	NOUN	post	NOUN
late	ADJECTIVE	o	NOUN	post	VERB
lateral	ADJECTIVE	ode	NOUN	posture	NOUN
league	NOUN	oeuvre	NOUN	pot	NOUN
legacy	NOUN	olein	NOUN	pound	NOUN
lemma	NOUN	ology	NOUN	pound	VERB

Stem	POS	Stem	POS	Stem	POS
prise	VERB	sire	VERB	test	VERB
pro	NOUN	sis	NOUN	thane	NOUN
prove	VERB	site	NOUN	theca	NOUN
ptosis	NOUN	size	NOUN	there	NOUN
pula	NOUN	sol	NOUN	therm	NOUN
pulse	NOUN	sole	NOUN	tic	NOUN
pus	NOUN	sole	VERB	tide	NOUN
quat	NOUN	solute	NOUN	tile	NOUN
quest	NOUN	solve	VERB	time	NOUN
quit	VERB	som	NOUN	tin	NOUN
r	NOUN	son	NOUN	tine	NOUN
range	VERB	sorb	VERB	tint	NOUN
ranger	NOUN	sort	NOUN	tint	VERB
rate	VERB	sort	VERB	tire	NOUN
re	NOUN	sperm	NOUN	tire	VERB
rectory	NOUN	spy	VERB	tom	NOUN
relative	NOUN	stable	NOUN	tome	NOUN
rest	NOUN	stall	VERB	ton	NOUN
rest	VERB	stance	NOUN	tonus	NOUN
ride	VERB	state	NOUN	tope	NOUN
rive	VERB	sterol	NOUN	tor	NOUN
rogation	NOUN	still	VERB	tract	NOUN
rum	NOUN	stole	NOUN	tractile	ADJECTIVE
s	NOUN	strain	VERB	tribe	NOUN
sail	VERB	sty	NOUN	tribute	NOUN
say	NOUN	style	NOUN	trope	NOUN
say	VERB	style	VERB	trophy	NOUN
scant	VERB	sue	VERB	uric	ADJECTIVE
scend	VERB	suit	NOUN	valve	NOUN
scent	NOUN	surd	NOUN	vamp	VERB
scopal	ADJECTIVE	surd	ADJECTIVE	vantage	NOUN
scope	NOUN	t	NOUN	vender	NOUN
scribe	VERB	tack	NOUN	vent	VERB
script	NOUN	tack	VERB	vent	NOUN
script	VERB	tact	NOUN	venue	NOUN
sec	NOUN	taint	VERB	verb	NOUN
sect	NOUN	tan	NOUN	verge	VERB
sense	NOUN	tax	NOUN	verse	NOUN
sent	NOUN	taxis	NOUN	verse	VERB
sent	ADJECTIVE	te	NOUN	vest	VERB
sept	NOUN	tech	NOUN	vet	NOUN
serine	NOUN	tee	NOUN	vise	NOUN
serve	NOUN	tee	VERB	visible	ADJECTIVE
serve	VERB	temper	NOUN	visor	NOUN
shop	NOUN	temper	VERB	void	VERB
sib	NOUN	tempt	VERB	void	ADJECTIVE
side	NOUN	tend	VERB	vote	VERB
side	VERB	tense	NOUN	y	NOUN
signor	NOUN	tense	ADJECTIVE	zeugma	NOUN
sin	NOUN	tensor	NOUN	zoic	ADJECTIVE
sine	NOUN	tent	NOUN		
sire	NOUN	tent	VERB		

Appendix 39

Section from initial concatenation analysis results

Original word	1st. component	Middle component	Final component
adage	ad		age
adapt	ad		apt
adaptability	ad	apt	ability
adaptable	ad	apt	able
adaption	ad	apt	ion
adaxial	ad		axial
adaxially	ad		axially
addition	ad	dit	ion
address	ad		dress
addressable	ad	dress	able
addressed	ad		dressed
adduct	ad		duct
adduction	ad	duct	ion
adequate	ad		equate
adhere	ad		here
adherent	ad	he	rent
adjoin	ad		join
adjudge	ad		judge
adjunction	ad		junction
adjust	ad		just
adjustable	ad	just	able
adjutant	ad	jut	ant
adman	ad		man
admass	ad		mass
admeasure	ad		measure
administer	ad		minister
administration	ad		ministration
admiration	ad	mi	ration
admire	ad		mire
admired	ad	mi	red
admission	ad	miss	ion
admission	ad		mission
admissive	ad		missive
admittable	ad	mitt	able
admix	ad		mix
admixture	ad		mixture
adnoun	ad		noun
adoptable	ad	opt	able
adoption	ad	opt	ion
adoration	ad		oration
adore	ad		ore
adrift	ad		rift
adscript	ad		script
adsorb	ad		sorb
adsorbable	ad	sorb	able
adsorption	ad		sorption
adulthood	ad	ult	hood
advancement	ad	van	cement

Original word	1st. component	Middle component	Final component
advent	ad		vent
adventure	ad		venture
adventuresome	ad	venture	some
adverb	ad		verb
adverse	ad		verse
advice	ad		vice
advisable	ad	vi	sable
advisee	ad	vi	see
advowson	ad	vow	son

Appendix 40

Concatenation first component stoplist

ace	act	ad	ado	aft
after	airs	all	alter	amp
ant	anti	arc	arch	art
as	ash	ask	ass	audit
auto	ax	back	bad	bag
ban	bar	barb	bash	bat
be	beg	best	bet	bill
bin	bit	blab	bob	bolo
bomb	boo	bore	bud	bug
bus	but	butt	by	cab
can	cant	cap	car	cart
cast	cat	cent	champ	chap
chic	chin	clan	clot	con
cop	corn	count	counter	cow
cows	cross	cry	cup	cur
dam	deter	din	dip	disc
do	dog	don	dot	down
drag	dry	due	eggs	end
enter	era	even	ever	extra
eyes	fan	far	fat	fig
flu	foe	form	formal	found
fun	fur	gal	gem	gig
glut	go	god	gram	grand
grim	grin	habit	habitat	halo
ham	harp	hat	hem	hero
hex	hi	hip	hot	hum
imp	in	inter	jab	jar
kit	lam	lap	lat	leg
less	let	lit	lob	log
lust	ma	maid	man	mar
marsh	mass	mat	men	mid
min	miss	mist	mix	mode
moo	muff	mull	neo	no
none	not	now	off	on
os	out	over	overt	ox

pa	pad	pale	pall	pan
pant	pap	par	pare	part
pass	past	pat	path	pen
pet	phone	photo	pie	pig
pill	plan	plat	plum	pole
poll	pop	port	post	pot
pro	prop	proto	prove	pseudo
puff	pun	pup	put	quasi
rabbi	radio	ram	rap	rat
ray	real	reap	red	rein
rest	rev	rhino	rig	rob
rot	saga	sap	scar	sea
sec	sect	see	sept	serge
set	sex	shy	sic	side
sigh	sign	sin	sing	sir
sis	slit	so	son	span
spic	stem	step	steps	stereo
stub	sub	sum	sun	super
supra	surge	tan	tar	tart
tat	taut	tax	tea	tee
tempo	ten	term	thin	thresh
through	tie	tin	tip	tit
ton	too	top	trim	trip
troops	tub	ultra	under	up
verb	vie	vow	wag	war
warp	wee	weir	whir	whit
win	wit	woo	woods	works
writ	zoo			

Appendix 41

Concatenation last component startlist

about	ache	acre	acting	after
afternoon	agent	air	aircraft	all
along	ambitious	angel	angelic	antibody
apple	arch	argument	arm	around
arrow	ash	asset	away	awe
axe	baby	back	backer	bacteria
bag	bait	bake	baked	bald
ball	band	bang	bank	bar
bare	bark	barn	base	basin
basket	bat	bath	bathe	bay
beak	beam	bean	bear	beard
bearer	bearing	beat	bedding	bee
beetle	before	being	bell	belly
below	belt	bench	bend	berg
berry	bill	bin	bind	binder
binding	bird	birth	bit	bite
black	blade	blast	bleed	blend

blind	block	blood	blot	blow
blower	blown	board	boarding	boat
bodice	body	boil	boiler	bold
bolt	bomb	bone	bonnet	book
booth	bore	born	boss	bottle
bottom	bound	bow	bowl	box
boy	brain	brake	brand	bread
breadth	break	breaking	breast	brick
bridge	brier	broken	broker	brow
brown	brush	buck	buckle	bud
bug	build	builder	building	bulb
bum	bump	burn	burner	burning
burnt	burst	bus	bush	butt
button	cab	cage	cake	call
can	candle	cane	cannon	cap
car	card	care	cart	carving
case	cast	castle	cat	catcher
cater	cellar	centrifuge	chair	chamber
chart	chase	chat	check	cheese
chick	child	choke	chop	chuck
clad	claim	clap	clasp	claw
clay	clean	clip	cloth	clothes
cloud	club	coach	coast	coat
cock	code	color	colored	colour
comb	comer	coming	cone	coop
cord	core	corn	corner	cotton
count	counter	course	court	cover
crack	cracker	craft	craftsman	cream
creeper	crest	crib	crop	cross
crossed	crossing	crow	crunch	cuff
cup	cushion	cut	cute	cycle
cyclist	dam	damp	dance	dancer
dash	day	days	dealer	decency
deck	deer	desk	devil	dew
dial	dig	dine	disc	disk
dive	dock	dog	door	dose
dough	dove	down	doze	dragon
draper	draw	drawn	dream	dress
dresser	dried	drift	driver	drop
drum	dust	eagle	ear	east
eastern	eastward	easy	edge	edit
eye	eyed	face	faced	fair
fall	fallen	fast	fat	father
fault	feast	feather	feed	feeder
felicity	fellow	field	fielder	fight
fighter	file	fill	film	final
finding	finger	fingered	fire	first
fish	fisher	fishing	fitting	flake
flap	flash	flask	flesh	flight
flint	float	flood	flour	flow

flower	fly	flyer	flying	foil
fold	folk	foot	force	forest
forge	fork	form	forte	forth
forward	found	founding	fowl	frame
free	freight	friend	frog	front
fruit	full	fund	gallant	game
gap	gas	gate	gather	gay
gaze	gear	gig	girl	giver
giving	glass	glory	glove	going
good	gorge	gown	grade	grain
graph	grass	grate	grave	green
grip	grocer	groom	ground	grown
growth	grudge	guard	guest	guide
guilt	gull	gun	gut	hack
hair	half	hall	hammer	hand
handle	happy	hard	hardy	harp
hat	hatch	hawk	head	headed
heap	heart	held	hell	hen
herb	herd	hide	hike	hill
hive	hog	hold	holder	holding
hole	hook	hop	hopper	horn
horse	hound	house	hunt	hunter
husband	incense	ionic	iron	jacket
jam	jar	jaw	jet	job
journalism	journalist	joy	keep	keeping
kerchief	kettle	kick	kill	killer
knife	knight	knob	knot	lace
laced	ladder	lady	lag	lamp
land	language	lap	lash	lasting
laugh	law	lay	layer	laying
lead	leader	leaf	leech	leg
legged	length	letter	lever	lick
lid	lie	life	lift	light
lighted	lighting	line	liner	link
lip	lipped	list	load	loaf
lobe	location	lock	locker	loft
long	loom	loose	lord	lore
louse	love	lover	luck	lust
luster	lustre	ma'am	made	maid
maiden	mail	maker	making	man
mane	march	mare	mark	market
mask	mass	mast	master	mat
match	meal	meat	meet	melon
metal	meter	milk	mill	mind
minded	mint	mistress	mobile	mold
month	moon	mop	moss	moth
mother	mould	mount	mouse	mouth
mow	much	muff	nail	name
naught	neck	nephew	net	niece
night	nip	nose	nosed	numerical

nurse	nut	oat	off	only
ounce	over	owner	pack	packing
pad	paint	pan	paper	parent
park	part	past	paste	pat
patch	path	pea	penny	people
perch	person	phone	phrase	pick
piece	pigeon	pile	pin	pipe
piper	pit	place	plain	plan
plane	plank	plant	plaster	plate
play	player	plow	plug	plum
pocket	point	poise	poke	pole
poll	pond	pool	port	position
positive	post	powder	power	press
prick	print	proof	prop	puff
pull	puncher	puppy	purse	quake
quarter	quest	race	radish	rag
rail	raise	rake	rat	rate
reach	read	reader	ready	reel
regal	rein	rending	rib	ride
rider	rig	rigger	right	road
robber	robe	rock	rocket	rod
roll	roof	room	roost	root
round	royal	royalty	rug	run
runner	running	rush	sack	saddle
safe	sake	sale	same	sand
sap	sauce	saver	saving	saw
scarf	school	scope	score	screen
seal	seat	seed	seeker	seer
sense	sensible	setting	shackle	shade
shadow	shaft	shake	shaking	shape
share	sharp	shave	sheet	shelf
shell	shield	shift	shine	shirt
shit	shod	shoe	shoot	shooter
shooting	shop	shore	shot	show
shower	sick	side	sight	signal
sill	silver	sit	site	sitting
skin	skirt	slaughter	sleeve	slide
slip	snail	snake	snap	snuffer
sock	soiled	song	sore	space
span	speak	speaker	speck	speed
spell	spike	spirited	spit	splitting
spoken	spoon	sport	spot	spout
spread	spring	spur	square	stack
staff	stain	stake	stalk	stamp
stand	standing	star	start	station
stay	stead	steak	stem	step
stern	stick	sticking	stitch	stock
stocking	stone	stool	stop	store
storm	stove	strain	strap	straw
streak	stream	stretch	stretched	stricken

strife	string	strip	stripe	stroke
strong	strung	stuff	style	sucker
suds	suit	sum	surf	sward
sweep	sweeping	sweet	swing	swipe
sword	tag	tail	take	tale
talk	tap	tape	teacher	telling
tender	terrier	therapy	think	thinker
thinking	thirsty	thorn	thread	throat
throb	through	tick	tide	tiger
tight	time	timer	times	tip
tit	toe	tongue	tooth	top
torch	total	totter	towel	tower
town	track	trap	tree	trot
truck	tub	tube	tuft	under
up	vendor	vine	virus	wad
wag	wagon	waist	waiter	walk
wall	warming	wart	wash	washing
watch	watcher	water	wave	wax
waxen	way	ways	wealth	wear
weed	week	weight	weir	weld
well	west	westerly	western	westward
whack	wheat	wheel	while	whip
whisk	whistle	white	wide	width
wife	wig	will	wind	window
wing	wings	wink	winner	winning
wire	wit	withal	witness	woman
wood	woods	wool	word	work
worker	working	works	world	worm
worn	worth	worthy	woven	wrap
wreck	wrestle	write	writer	writing
yard				

Appendix 42

Words starting with "non-" and "un-" which are not antonymous prefixations

FROM nonaginta = ninety

nonagenarian

FROM nonus = ninth

nones

FROM no

none, nonesuch, nonetheless, nonsuch

MISLEADING ANTONYMOUS PREFIX non-

nonage, nonaged, nonallele, nonchalance, nonchalant, nonchalantly, nonplus, nonplused, nonplussed,

UNCERTAIN non-

nonagon, nonce, noncom, nonuple

PREFIX under

under, underachieve, underachievement, underachiever, underact, underactive, underage, underarm, underbelly, underbid, underbodice, underbody, underboss, underbred, underbrush, undercarriage, undercharge, underclass, underclassman, underclothed, underclothes, underclothing, undercoat, undercoated, undercover, undercover agent, undercover operation, undercover work, undercurrent, undercut, underdevelop, underdeveloped, underdevelopment, underdog, underdone, underdrawers, underdress, underdressed, undereducated, underemployed, underestimate, underestimation, underevaluation, underexpose, underexposure, underfed, underfelt, underfoot, underframe, underfur, undergarment, undergird, undergo, undergrad, undergraduate, underground, underground press, undergrow, undergrowth, underhand, underhanded, underhandedly, underhung, underlay, underlayment, underlie, underline, underling, underlip, underlying, undermanned, undermentioned, undermine, underneath, undernourish, undernourished, undernourishment, underpants, underpart, underpass, underpay, underpayment, underperform, underperformer, underpin, underplay, underpopulated, underprice, underprivileged, underproduce, underproduction, underquote, underrate, underrating, underreckoning, underscore, undersea, underseal, undersealed, undersecretary, undersell, underseller, undersexed, undershirt, undershoot, undershot, undershrub, underside, undersign, undersize, undersized, underskirt, underslung, undersoil, underspend, understaffed, understand, understandability, understandable, understandably, understanding, understandingly, understate, understated, understatement, understock, understood, understructure, understudy, undersurface, undertake, undertaker, undertaking, undertide, undertone, undertow, undervaluation, undervalue, underwater, underwater archaeology, underwater archeology, underwater diver, underway, underwear, underweight, underwing, underwood, underworld, underwrite, underwriter

BUT ANTONYMOUS PREFIX un- before der

underivative, underived

PREFIX undula "wave"

undulant, undulant fever, undulate, undulation, undulatory, undulatory theory

PREFIX uni-

unicameral, unicameral script, unicellular, unicorn, unicorn , root, unicuspid, unicycle, unicyclist, unidimensional, unidirectional, unifacial, unification, unified, unifilar,

unifoliate, uniform, uniform resource locator, uniformed, uniformise, uniformity, uniformize, uniformly, uniformness, unify, unifying, unilateral, unilateral contract, unilateral descent, unilateral paralysis, unilateralism, unilateralist, unilaterally, unimodal, uninominal, uninominal system, uninominal voting system, uninucleate, uniocular , dichromat, union, union card, union member, union representative, union shop, union suit, unionisation, unionise, unionised, unionism, unionist, unionization, unionize, unionized, uniovular, uniovulate, uniparous, unipolar, unipolar , depression, unique, uniquely, uniqueness, unisex, unisexual, unison, unit, unit cell, unit character, unit cost, unit investment , trust, unit matrix, unit of ammunition, unit of measurement, unit of , time, unit of viscosity, unit trust, unitard, unitary, unite, united, unitedly, uniting, unitisation, unitise, unitization, unitize, unity, univalent, univalve, universal, universal agent, universal , donor, universal gas constant, universal gravitational constant, universal , joint, universal proposition, universal quantifier, universal resource locator, universal set, universal solvent, universal suffrage, universal time, universal veil, universalise, universalism, universalist, universalistic, universality, universalize, universally, universe, universe of , discourse, university, university extension, university student, univocal

BUT ANTONYMOUS PREFIX un- before i

unidentifiable, unidentified, unidentified flying object, unilluminated, unilluminating, unimaginable, unimaginably, unimaginative, unimaginatively, unimagined, unimpaired, unimpassioned, unimpeachable, unimpeachably, unimpeded, unimportance, unimportant, unimposing, unimpregnated, unimpressed, unimpressionable, unimpressive, unimpressively, unimprisoned, unimproved, unincorporated, unindustrialised, unindustrialized, uninebriated, uninfected, uninflected, uninfluenced, uninfluential, uninformative, uninformatively, uninformed, uninhabitable, uninhabited, uninhibited, uninitiate, uninitiated, uninjectable, uninjured, uninquiring, uninquisitive, uninspired, uninspiring, uninstructed, uninstructive, uninstructively, uninsurability, uninsurable, uninsured, unintegrated, unintelligent, unintelligently, unintelligibility, unintelligible, unintelligibly, unintended, unintentional, unintentionally, uninterested, uninteresting, uninterestingly, uninterestingness, uninterrupted, uninterruptedly, unintimidated, unintoxicated, unintrusive, uninventive, uninvited, uninvitedly, uninviting, uninvolved, unironed

PREFIX un- for uni before vowel

unanimity, unanimous, unanimously, unary, unary operation

PREFIX -unc "annoit"

unction, , unctuous, unctuously, unctuousness

PREFIX -ung "annoit"

unguent

PREFIX ungula "nail"

ungulate, ungulated

ATOMIC

uncle

NON-ANTONYMOUS PREFIX un-

until, unto

Appendix 43

Antonymous prefixation exceptions and counter-exceptions
(Whole word exceptions not shown)

Morpheme exceptions

under	undula	uni	unanim
unary	unct	ungula	infra
inner	inq	inb	inl
inm	inp	inr	inw
integr	intellect	intellig	inter
integument	intra	intro	inch
india	ink	ana	ante
antiqu	annoy	anoint	anomal
answer	anxious	any	andro
anb	anc	and	anf
ang	anj	ank	anl
anm	ann	anp	anq
anr	ans	antb	antc
antd	antf	antg	antj
antk	antl	antm	antn
antp	antq	antr	ants
antt	antv	antw	antx
anty	antz	anemo	angel
anger	angio	angle	angl
ango	angri	anguish	angular
anima	animal	animate	anim
ankle	annal	anneal	annelid
annex	annihilat	annual	annotat
announce	annunciat	anorec	anorex
antho	anthrop	aa	ae
ah	ai	ao	au
aw	ay	contrb	contrc
contrd	contrf	contrg	contrh
contrj	contrk	conrl	contrm
contrn	comtrp	contrq	contrr
contrs	contrt	contrv	contrw
contrx	contrz	contraa	contrae
contrai	contrao	contrau	countera
counterb	counterc	counterd	countere

counterf	counterg	counterh	counteri
counterj	counterk	counterl	counterm
countern	countero	counterp	counterq
counterr	counters	countert	counteru
counterv	counterw	counterx	countery
counterz			

Whole word counter-exceptions

unidentifiable	unidentified	unilluminated	unilluminating
unimaginable	unimaginably	unimaginative	unimaginatively
unimagined	unimpaired	unimpassioned	unimpeachable
unimpeachably	unimpeded	unimportance	unimportant
unimposing	unimpregnated	unimpressed	unimpressionable
unimpressive	unimpressively	unimprisoned	unimproved
unincorporated	unindustrialised	unindustrialized	uninebriated
uninfected	uninflected	uninfluenced	uninfluential
uninformative	uninformatively	uninformed	uninhabitable
uninhabited	uninhibited	uninitiate	uninitiated
uninjectable	uninjured	uninquiring	uninquisitive
uninspired	uninspiring	uninstructed	uninstructive
uninstructively	uninsurability	uninsurable	uninsured
unintegrated	unintelligent	unintelligently	unintelligibility
unintelligible	unintelligibly	unintended	unintentional
unintentionally	uninterested	uninteresting	uninterestingly
uninterestingness	uninterrupted	uninterruptedly	unintimidated
unintoxicated	unintrusive	uninventive	uninvited
uninvitedly	uninviting	uninvolved	unironed
interminable	interminably	intractability	intractable
intractableness	intractably	intransigence	intransigency
intransigent	intransitive	intransitively	intransitiveness
intransitivise	intransitivity	intransitivize	introuvable
anaemia	anaemic	anaerobe	anaerobic
anaerobiotic	anaesthesia	anaesthetic	anaesthetise
anaesthetist	anaesthetize	analphabet	analphabetic
analphabetism	anaphrodisia	anaphrodisiac	anapsid
anarchic	anarchical	anarchically	anarchism
anarchist	anarchistic	anarchy	anarthria
anaspid	antacid	antagonise	antagonism
antagonist	antagonistic	antagonistically	antagonize
antapex	arrhythmia	arrhythmic	arrhythmical
anomia	anomic	anomie	anomy
counterclockwise	counterintuitive	counterintuitively	

Morpheme counter-exceptions

underiv	analges	anti	aneur
antonym	anomal		

Appendix 44

1st. secondary suffix set as ordered by the optimal heuristic

ing	er	e	ed	al
ate	ation	ion	ic	on
ine	able	ent	ive	age
ight	ly	ble	ism	ter
tion	like	ness	ist	ity
th	ish	ology	ify	ng
ification	ingly	ally	ess	us
ful	ower	tor	tic	ck
ical	ise	ard	ough	ook
idity	y	ow	s	ch
ted	sh	t	an	ike
ility	ighted	ular	our	ative
ings	ound	ide	ting	um
atory	ogy	ize	te	own
ator	ette	ified	out	le
ment	istic	ack	ability	ip
lessness	ightly	ookie	inate	ated
ically	iveness	ail	ope	ologist
ram	ounding	ght	in	ome
n	eeder	ood	ark	ia

Appendix 45

Homonyms with POS variation: result samples

Homonym1	POS1	Homonym2	POS2	Relation type
100	NOUN	100	ADJECTIVE	DERIV
Burundi	NOUN	Burundi	ADJECTIVE	DERIV
Ghanian	ADJECTIVE	Ghanian	NOUN	DERIV
Mandaean	ADJECTIVE	Mandaean	NOUN	DERIV
Proterozoic	NOUN	proterozoic	ADJECTIVE	DERIV
Uniate	ADJECTIVE	Uniate	NOUN	DERIV
advance	NOUN	advance	ADJECTIVE	DERIV
amber	NOUN	amber	ADJECTIVE	DERIV
aphrodisiac	NOUN	aphrodisiac	ADJECTIVE	DERIV
audible	ADJECTIVE	audible	NOUN	DERIV
bag	NOUN	bag	VERB	DERIV
battle	VERB	battle	NOUN	DERIV
bias	VERB	bias	NOUN	ROOT
blank	VERB	blank	NOUN	DERIV
boil	NOUN	boil	VERB	DERIV
branch	VERB	branch	NOUN	DERIV
buckram	VERB	buckram	NOUN	DERIV
bypass	VERB	bypass	NOUN	DERIV
caramel	ADJECTIVE	caramel	NOUN	DERIV
censor	NOUN	censor	VERB	DERIV
cheat	NOUN	cheat	VERB	DERIV
claim	NOUN	claim	VERB	DERIV
cluck	VERB	cluck	NOUN	DERIV
compare	NOUN	compare	VERB	DERIV
cook	VERB	cook	NOUN	DERIV
crack	NOUN	crack	ADJECTIVE	DERIV
crosscut	NOUN	crosscut	VERB	DERIV
dab	VERB	dab	NOUN	DERIV
deictic	NOUN	deictic	ADJECTIVE	DERIV
dirt	NOUN	dirt	ADJECTIVE	DERIV
douche	NOUN	douche	VERB	DERIV
drum	NOUN	drum	VERB	DERIV
egress	NOUN	egress	VERB	DERIV
erotic	ADJECTIVE	erotic	NOUN	DERIV
fain	ADJECTIVE	fain	ADVERB	DERIV
ferret	NOUN	ferret	VERB	DERIV
flame	NOUN	flame	VERB	DERIV
flux	NOUN	flux	VERB	DERIV
frank	NOUN	frank	ADJECTIVE	DERIV
gag	NOUN	gag	VERB	DERIV
gibbet	NOUN	gibbet	VERB	DERIV
gown	NOUN	gown	VERB	DERIV
guard	VERB	guard	NOUN	DERIV
hatch	VERB	hatch	NOUN	DERIV
hinge	NOUN	hinge	VERB	DERIV
hotfoot	VERB	hotfoot	NOUN	DERIV
impact	VERB	impact	NOUN	DERIV

Homonym1	POS1	Homonym2	POS2	Relation type
interlock	VERB	interlock	NOUN	DERIV
jitterbug	VERB	jitterbug	NOUN	DERIV
kip	NOUN	kip	VERB	DERIV
last	ADVERB	last	ADJECTIVE	DERIV
lilliputian	NOUN	lilliputian	ADJECTIVE	DERIV
lurch	NOUN	lurch	VERB	DERIV
mass	VERB	mass	NOUN	ROOT
midland	ADJECTIVE	midland	NOUN	DERIV
molar	ADJECTIVE	molar	NOUN	DERIV
mug	VERB	mug	NOUN	DERIV
net	NOUN	net	ADJECTIVE	DERIV
off	ADVERB	off	ADJECTIVE	DERIV
outside	ADVERB	outside	ADJECTIVE	DERIV
palsy	NOUN	palsy	VERB	DERIV
pattern	NOUN	pattern	VERB	DERIV
philharmonic	NOUN	philharmonic	ADJECTIVE	DERIV
plain	ADJECTIVE	plain	ADVERB	DERIV
polish	VERB	polish	NOUN	DERIV
precis	VERB	precis	NOUN	DERIV
programme	NOUN	programme	VERB	DERIV
purport	NOUN	purport	VERB	DERIV
rabbit	VERB	rabbit	NOUN	DERIV
rebound	VERB	rebound	NOUN	DERIV
remote	ADJECTIVE	remote	NOUN	DERIV
revere	VERB	revere	NOUN	DERIV
roof	VERB	roof	NOUN	DERIV
sallow	ADJECTIVE	sallow	NOUN	DERIV
schmooze	NOUN	schmooze	VERB	DERIV
seat	NOUN	seat	VERB	DERIV
shame	VERB	shame	NOUN	DERIV
shuck	NOUN	shuck	VERB	DERIV
skid	VERB	skid	NOUN	DERIV
slum	VERB	slum	NOUN	DERIV
snow	NOUN	snow	VERB	DERIV
spar	VERB	spar	NOUN	DERIV
spree	VERB	spree	NOUN	DERIV
star	NOUN	star	ADJECTIVE	DERIV
store	VERB	store	NOUN	DERIV
submarine	VERB	submarine	NOUN	ROOT
suture	NOUN	suture	VERB	DERIV
take	VERB	take	NOUN	DERIV
tent	VERB	tent	NOUN	DERIV
thyroid	ADJECTIVE	thyroid	NOUN	DERIV
touch	NOUN	touch	VERB	DERIV
tricolor	ADJECTIVE	tricolor	NOUN	DERIV
twin	NOUN	twin	ADJECTIVE	DERIV
uplift	VERB	uplift	NOUN	DERIV
virgin	ADJECTIVE	virgin	NOUN	DERIV
wassail	VERB	wassail	NOUN	DERIV
white	VERB	white	NOUN	ROOT
wrestle	NOUN	wrestle	VERB	DERIV

Appendix 46

Secondary concatenation last component startlist

abed	act	age	ass	bed
by	chant	clerk	ease	end
fare	few	hip	hood	key
kind	lance	like	linger	mania
maniac	mate	men	mine	more
most	note	one	out	page
pen	pie	pike	pot	rack
ray	rest	ring	rope	rose
row	sail	say	script	see
set	shed	sing	size	sole
some	son	stall	still	story
sure	table	tack	tease	thing
tie	tone	train	tray	trip
wed	written			

Appendix 47

Secondary concatenation complementary first component stoplist

add	allot	check	clay	coin
coon	hinder	hub	lag	lug
moss	rag	rug	summer	tube

Appendix 48

Secondary concatenation analysis results (complete)

Original word	1st. component	Last component	Original word	1st. component	Last component
airfare	air	fare	egotrip	ego	trip
anymore	any	more	eightsome	eight	some
armrest	arm	rest	fadeout	fade	out
ballpen	ball	pen	fallout	fall	out
banknote	bank	note	farthermost	farther	most
bannerlike	banner	like	featherbed	feather	bed
bedrest	bed	rest	feverfew	fever	few
blackout	black	out	fieldfare	field	fare
bloodshed	blood	shed	fingerstall	finger	stall
blowout	blow	out	fivesome	five	some
bookend	book	end	flatbed	flat	bed
bookstall	book	stall	flatmate	flat	mate
bottommost	bottom	most	flowerbed	flower	bed
bowtie	bow	tie	flowerpot	flower	pot
breakout	break	out	foldout	fold	out
brownout	brown	out	footnote	foot	note
bullpen	bull	pen	footrest	foot	rest
bullring	bull	ring	footstall	foot	stall
bunkmate	bunk	mate	footsure	foot	sure
businessmen	business	men	forevermore	forever	more
buyout	buy	out	foursome	four	some
campmate	camp	mate	freelance	free	lance
chamberpot	chamber	pot	frontmost	front	most
childbed	child	bed	frontstall	front	stall
chimneypot	chimney	pot	furthermore	further	more
classmate	class	mate	furthermost	further	most
clearstory	clear	story	fusspot	fuss	pot
closeout	close	out	gainsay	gain	say
coatrack	coat	rack	gearset	gear	set
cocksure	cock	sure	geartrain	gear	train
coffeepot	coffee	pot	goldmine	gold	mine
cookout	cook	out	goodby	good	by
crackpot	crack	pot	gunslinger	gun	linger
cutout	cut	out	halftone	half	tone
daybed	day	bed	handout	hand	out
deathbed	death	bed	handrest	hand	rest
dimout	dim	out	handset	hand	set
dropout	drop	out	hangout	hang	out
dumbass	dumb	ass	hardtack	hard	tack
earring	ear	ring	hayrack	hay	rack
easternmost	eastern	most	headrest	head	rest
eastmost	east	most	headsail	head	sail
egomania	ego	mania	headset	head	set
egomaniac	ego	maniac	headstall	head	stall

Original word	1st. component	Last component	Original word	1st. component	Last component
hearsay	hear	say	playscript	play	script
heartsease	heart	ease	plaything	play	thing
heavyset	heavy	set	porkpie	pork	pie
hedgerow	hedge	row	printout	print	out
helpmate	help	mate	pullout	pull	out
hereby	here	by	quickset	quick	set
hideout	hide	out	readout	read	out
hitchrack	hitch	rack	rearmost	rear	most
holdout	hold	out	rightmost	right	most
homepage	home	page	riverbed	river	bed
honeypot	honey	pot	roadbed	road	bed
housemate	house	mate	rockrose	rock	rose
humankind	human	kind	roommate	room	mate
icetray	ice	tray	rosehip	rose	hip
inkpot	ink	pot	roundtable	round	table
innermost	inner	most	salesclerk	sale	clerk
innersole	inner	sole	saucepot	sauce	pot
jampot	jam	pot	schoolmate	school	mate
keynote	key	note	seedbed	seed	bed
knockout	knock	out	sellout	sell	out
latchkey	latch	key	sevensome	seven	some
layby	lay	by	shakeout	shake	out
layout	lay	out	shipmate	ship	mate
leftmost	left	most	shootout	shoot	out
lifesize	life	size	shutout	shut	out
linemen	line	men	sickbed	sick	bed
lockout	lock	out	sightsee	sight	see
lockring	lock	ring	sightsing	sight	sing
lookout	look	out	sixsome	six	some
lowermost	lower	most	skysail	sky	sail
lowset	low	set	slugabed	slug	abed
mainsail	main	sail	someone	some	one
maniclike	manic	like	southernmost	southern	most
messmate	mess	mate	southmost	south	most
middlemost	middle	most	stablemate	stable	mate
mindset	mind	set	stakeout	stake	out
monkshood	monk	hood	stalemate	stale	mate
mudslinger	mud	linger	standby	stand	by
nearby	near	by	standstill	stand	still
necktie	neck	tie	staysail	stay	sail
nevermore	never	more	stingray	sting	ray
newlywed	newly	wed	stinkpot	stink	pot
northernmost	northern	most	stockpot	stock	pot
northmost	north	most	streambed	stream	bed
outermost	outer	most	strikeout	strike	out
plainchant	plain	chant	striptease	strip	tease
playact	play	act	suchlike	such	like
playmate	play	mate	tablemate	table	mate
playpen	play	pen	takeout	take	out

Original word	1st. component	Last component	Original word	1st. component	Last component
teammate	team	mate	typescript	type	script
teenage	teen	age	typeset	type	set
thereby	there	by	uppermost	upper	most
thickset	thick	set	uttermost	utter	most
thoroughfare	thorough	fare	walkout	walk	out
threesome	three	some	washout	wash	out
thumbstall	thumb	stall	watershed	water	shed
thumbtack	thumb	tack	webpage	web	page
ticktack	tick	tack	weekend	week	end
tightrope	tight	rope	westernmost	western	most
timetable	time	table	westmost	west	most
toastrack	toast	rack	whiteout	white	out
toolshed	tool	shed	whoreson	whore	son
towrope	tow	rope	wipeout	wipe	out
tryout	try	out	womankind	woman	kind
turnkey	turn	key	woodshed	wood	shed
turnout	turn	out	workmate	work	mate
turnpike	turn	pike	workout	work	out
turntable	turn	table	worktable	work	table
twosome	two	some			

Appendix 49

Irregular prefixes with sample instances

Footprint	Prefix name	Character sequence to delete	Character sequence to insert	Sample instances
abb	abba	abb		abbacy, abbatial, abbe, abbess, abbey
abb	ad	ab		abbreviate, abbreviated, abbreviation, abbreviator
absc	ab	abs		abscess, abscessed, abscond, absconder, abscondment
abst	ab	abs		abstract, abstracted, abstractedly, abstractedness, abstracter
ab	ab	ab		abarticulation, abaxial, abaxially, abdicable, abdicate
ab	a	a		aback, abase, abasement, abash, abashed
ab	a	ab		abaft
ab	a1	a		abnormal, abnormalcy
ab	ad	a		abandon, abandoned, abandonment, abatable, abate
acc	ad	ac		accede, accelerando, accelerate, accelerated, acceleration
acc	a	ac		accurse, accursed, accurst
ach	ad	a		achieve
acq	ad	ac		acquaint, acquaintance, acquaintanceship, acquainted, acquiesce
acri	acri	acri		acrid, acridid, acrimony
adolesc	adolesc	adolesc		adolesce, adolescence, adolescent
adult	adult	adult		adult, adulterant, adulterate, adulterated, adulterating
ad	ad	ad		adaxial, adaxially, addict, addicted, addiction
ad	a	a		ado, adrift, adamance, adamant, adamantine
aff	ad	af		affability, affable, affableness, affably, affair

Footprint	Prefix name	Character sequence to delete	Character sequence to insert	Sample instances
aff	a	af		afford, affordable, affright, affront
aff	ex	af		affray
agg	ad	ag		agglomerate, agglomerated, agglomeration, agglomerative, agglomerator
ali	ali	ali		alias, alibi, alien
allo	allo	allo		alloantibody, allochronic, allochthonous, allogeneic, allograph
all	allo	all		allegoric, allegorical, allegorically, allegorise, allegoriser
all	ad	al		alla, allargando, alleviant, alleviate, alleviated
all	a	al		allay, allayer
alter	altr	alter		alter, altercate, alternate, alternative
alti	alt	alti		altimeter, altissimo, altitude, altitudinous
alto	alt	alto		alto, altocumulus, altostratus
altr	altr	altr		altruism
al	all	al		almighty, already, alright, also, altogether
amm	ad	am		ammo, ammunition
amm	amp	am		ammeter
am	am	am		amateur, amative, amatory, amenity, amiable
am	ad	a		ameliorate, amenable, amerce, amerciable, amort
am	ex	a		amend, amends
ana	ana	ana		anabiosis, anabiotic, anabolic, anabolism, anachronic
ancest	ante	an		ancestor
ancient	ante	ancient		ancient
andro	andro	andro		androecium, androgen, androgenesis, androgenetic, androgenic
andr	andro	andr		andradite, andrena, andrenid, andryala
anemo	anemo	anemo		anemone, anemographic, anemography, anemometer, anemometric
ang	ank	ang		angst, anger, angry
anni	ann	anni		anniversary
annu	ann	annu		annual, annuitant, annuity, annum
annu	annu	annul		annular, annulate, annulet, annulus
ann	ad	an		annotate, announce, annul, annulment, annunciate
ano	ano	ano		anorectal, anorectic, anorexia, anorexic, anorexigenic
ante	ante	ante		antebellum, antecede, antecedence, antecedency, antecedent
anth	antho	anth		anthesis
antho	antho	antho		anthologise, anthologist, anthologize, anthology, anthophagous
antiqu	antiqu	antiqu		antiquary, antiquarian, antiquate, antiquated, antique
anti	anti	anti		antiacid, antiadrenergic, antiaircraft, antialiasing, antianxiety
ant	anti	ant		antacid, antagonise, antagonism, antagonist, antagonistic
anx	ank	anxi		anxiety, anxious
an	a	a		anew
an	a	an		another, answer, any
an	ana	an		anchorite, anion, anionic, anodal, anode
an	a1	an		anaemia, anaesthetise, anaesthetist, analbuminemia, analgesia
aperi	aperi	aperi		aperient, aperiodic, aperitif
apert	aperi	apert		aperture
aphro	aphro	aphro		aphrodisia, aphrodisiac, aphrodisiacal

Footprint	Prefix name	Character sequence to delete	Character sequence to insert	Sample instances
aph	apo	ap		aphaeresis, aphaeretic, aphelion, apheresis, apheretic
api	api	api		apicultural, apiculture, apiculturist, apivorous
app	ad	ap		apparatus, apparel, apparency, apparent, apparition
ap	a	a		apiece
archi	arch	archi		archidiaconal, archidiaconate, archiepiscopal
arch	arch	arch		archangel, archangelic, archbishop, archbishopric, archdeacon
arc	arc	arc		arccos, arccosecant, arccosine, arccotangent, arcdegree
arr	ad	ar		arraign, arraignment, arrange, arranged, arrangement
arr	err	arr		arrant
ass	ad	as		assail, assailability, assailable, assailant, assault
ass	ex	as		assay, assayer
ast	ex	a		astonied, astonish, astound
as	ad	a		ascend, ascent, ascertain, ascribe, aspect
ato	ad	at		atone
att	ad	at		attach, attachable, attache, attached, attachment
att	apt	att		attitude, attitudinal, attitudinise, attitudinize
av	ab	a		averse, avert
av	ad	a		avail, avenue, avocation
av	ex	a		avoid
a	a	a		acknowledge, afar, afeard, afield, afire
a	a1	a		acarpellous, acarpelous, acarpous, acephalia, acephalism
be	be	be		becalm, becharm, becloud, become, bedamn
cath	cata	cat		catharsis, cathartic, cathartid, cathect, cathectic
cat	cata	cat		catechesis, catechetic, catechetical, catechise, catechism
cogn	con	cog		cognomen
coll	con	col		collaborate, collaboration, collaborationism, collaborationist, collaborative
coll	col	coll		collage, collagen, collagenase, collagenic, collagenous
coll	coll	coll		collar, collarbone, collared, collarless, collet
coll	coll1	coll		collard, collards
coll	coll2	coll		collier, colliery
coll	coll3	coll		collywobbles
comb	con	com		combat, combatant, combative, combatively, combativeness
comme	comme	comme		comme
comm	con	com		command, commandant, commandeer, commander, commandership
comm	cop	comm		comma
comm	com	comm		commedia
compt	contra	compt		comptroller, comptrollership
comp	con	com		compact, compaction, compactly, compactness, companion
contra	contra	contra		contraband, contrabandist, contrabass, contrabassoon, contraception
contra	con	con		contract, contractable, contracted, contractile, contractility
contre	contra	contre		contredanse, contretemps, control, controllable, controlled
contr	contra	contro		controversial, controversialist, controversially, controversy, controvert
contr	contra	contr		contrast, contrasting, contrastingly, contrastive, contrasty

Footprint	Prefix name	Character sequence to delete	Character sequence to insert	Sample instances
con	cone	con		cone, coneflower, conelike, conic, conical
con	con	con		concatenate, concatenation, concave, concavely, concaveness
con	con	con		congelation, congenator, congener, congeneric, congenerical
con	con	con		consume, consumer, consumerism, consuming, consummate
corr	con	cor		correct, correctable, corrected, correction, correctional
corr	corr	corr		corridor
dead	die	dead		dead, deadbeat, deadbolt, deaden, deadened
death	die	death		death, deathbed, deathblow, deathless, deathlike
dea	dia	dea		deacon, deaconess
dea	deka	dea		dean, deanery, deanship
deb	deb	deb		debenture, debit, debitor, debt, debtor
deca	dec	deca		decade, decagon, decagram, decahedron, decaliter
dece	dec	dece		decennary, decennium
deci	dec	deci		decibel, decigram, deciliter, decilitre, decimal
deco	deco	deco		deco, decor, decorate, decorated, decoration
dec	deco	dec		decency, decent, decently
deed	deed	deed		deed, deedbox, deeds
dei	dei	dei		deific, deification, deify, deism, deist
del	del	del		delete, deleterious, deletion, delible
deka	deka	deka		dekagram, dekaliter, dekalitre, dekameter, dekametre
dema	dem	dema		demagog, demagogic, demagogical, demagogue, demagoguery
demi	demi	demi		demiglace, demigod, demimondaine, demimonde, demisemiquaver
demon	demon	demon		demon, demonetisation, demoniac, demoniacal, demoniacally
demo	dem	demo		democracy, democrat, democratic, democratically, democratisation
dendr	dendr	dendr		dendriform, dendrite, dendritic, dendrobium, dendroid
denti	denti	denti		denticle, denticulate, dentifrice, dentin, dentine
dent	denti	dent		dental, dentate, denture, denturist
dermati	derm	dermati		dermatitis
dermato	derm	dermato		dermatoglyphic, dermatoglyphics, dermatologic, dermatological, dermatologist
derm	derm	derm		derma, dermabrasion, dermal, dermic, dermis
desk	disco	desk		desk, deskbound, deskman, desktop
despot	despot	despot		despot, despotic, despotical, despotism
des	dis	des		dessert, dessertspoon, dessertspoonful, deshabille
deterior	deterior	deterior		deteriorate, deterioration
deuc	deu	deuc		deuce, deuced, deucedly
deuter	deuter	deuter		deuteranopia, deuteranopic, deuterium, deuteron
dexter	dextro	dexter		dexter, dexterity, dexterous, dexterously
dextro	dextro	dextro		dextral, dextrality, dextrin, dextroamphetamine, dextrocardia
de	de	de		decipher, decipherable, decipherably, deciphered, decipherer
de	de	de		defraud, defrauder, defray, defrayal, defrayment
de	de	de		depredation, depress, depressant, depressed, depressing
de	de	de		dehydroretinol, demineralise, demode, demodulate, demulcent
de	dia	de		devil, devilfish, devilise, devilish, devilishly

Footprint	Prefix name	Character sequence to delete	Character sequence to insert	Sample instances
dia	dia	di		diamante, diamantine, diamond
dia	di	di		diacetylmorphine, diapsid, diarchy, diazotize, diazoxide
dia	dia	dia		diabatic, diabetes, diabetic, diabolatry, diabolise
die	dia	di		dieresis
diff	dis	dif		differ, differentia, difficult, diffident, difflugia
dig	dis	di		digest, digestive, digress
dil	dis	di		dilapidate, dilate, diligent, diluent, dilute
dim	dis	di		dimension
dim	de	di		diminish, diminuendo, diminution, diminutive
dio	dia	di		diocesan, diocese, diorama
dir	dis	di		direct, directive, directory, dirigible
disc	disco	disc		disc, disciform, disclike, disco, discography
dish	disco	dish		dish, dishcloth, dished, dishful, dishpan
disk	disco	disk		disk, diskette, disklike
dis	dis	dis		disappoint, disappointed, disappointedly, disappointing, disappointingly
dis	dis	dis		disembowel, disentangler, disfluency, disgruntled, disparage
dis	di1	dis		dismal, dismally, dismay, distrain
dis	dis	di		dispersal, disperse, dispersed, dispersion, dispersive
dis	di	di		disyllabic, disyllable
diu	dia	di		diuresis, diuretic
div	dis	di		diverge, divers, diverse, divert, diverticulosis
di	di	di		dibrach, dibranch, dibranchiate, dibucaine, dicamptodon
di	di1	di		dial, diary, diet, dietetic, dietitian
ecclesi	ecclesi	ecclesi		ecclesiastic, ecclesiology
ecc	ex	ec		eccentric
echino	echino	echino		echinocactus, echinococcosis, echinococcus, echinoderm, echinus
echo	echo	echo		echocardiogram, echocardiograph, echocardiography, echoencephalogram, echoencephalograph
eco	eco	eco		ecobabble, ecology, econometric, econometrist, economy
ecto	ecto	ecto		ectoblast, ectoderm, ectodermic, ectomorph, ectomorphy
ecto	ec	ec		ectopia
ecu	eco	ecu		ecumenic, ecumenism
ec	ec	ec		ecchymosis, eccrine, eccyesis, ecdysiast, ecdysis
eff	ex	ef		efface, effect, effeminate, effeminise, efferent
ell	en	el		ellipse, ellipsis, ellipsoid, elliptic
emb	en	em		embalm, embank, embargo, embark, embarrass
emp	en	em		empale, empanel, empathy, empennage, emperor
end	endo	end		endameba, endemical, endemism, endergonic, endemic
end	en	en		endaemonism, endanger, endangered, endangerment, endear
eno	eno	eno		enologist, enology, enophile, enosis
entero	entero	entero		enterobacteria, enterobiasis, enteroceptor, enterokinase, enterolith
enter	enter	enter		enterprise, enterpriser, enterprising, enterprisingly, enterprisingness
enter	entero	enter		enteral, enteric, enterics, enteritis
entomo	entomo	entomo		entomion, entomologic, entomological,

Footprint	Prefix name	Character sequence to delete	Character sequence to insert	Sample instances
				entomologist, entomology
ento	ento	ento		entoblast, entoderm, entoparasite, entopic, entoproct
entre	inter	entre		entr'acte, entrecote, entree, entremets, entrepot
ent	en	en		entablature, entail, entailment, entangle, entangled
enu	ex	e		enucleate, enucleation, enumerable, enumerate, enumeration
en	en	en		enable, enabling, enact, enactment, enamor
en	ex	e		enate, enatic, enation, enounce
epan	epan	epan		epanalepsis, epanaphora, epanodos, epanorthosis
epaul	epaul	epaul		epaulet, epaulette, epauliere
eph	epi	ep		ephedra, ephedrine, ephemera, ephemeral, ephemerality
epi	epi	epi		epicalyx, epicanthic, epicanthus, epicardia, epicardium
epi	ex	e		epilate, epilation
ep	epi	ep		ependyma, epenthesis, epenthetic, epergne, eponym
es		e		escalade, escalate, escallop, escargot, escarole
es	ex	es		escape, escapade, escheat, escort, esplanade
eu	eu	eu		eubacteria, eubacterium, eucalypt, eucalyptus, euclidean
ev	eu	ev		evaporate, evaporite, evaporometer, evangel
exe	ex	ex	s	execrable, execrate, execration, executability, executable
exe	ex	ex		exenterate, exenteration, exercise, exerciser, exercising
exig	ex	exi	a	exigency, exigent, exiguity, exiguous
exi	ex	ex	s	exile, exilic, exist, existence, existent
exi	ex	ex		exit
exo	exo	exo		exobiology, exocarp, exocentric, exocrine, exoderm
exo	ex	ex	h	exode, exodus, exorcise, exorcism, exorcist
exo	ex	ex		exomphalos, exonerate, exonerated, exoneration, exonerative
exp	ex	ex	s	expect, expectable, expectancy, expectant, expectantly
exp	ex	ex		expat, expatiate, expatiation, expatriate, expatriation
exter	exter	exter		exterior, exteriorisation, exteriorise, exteriorization, exteriorize
extra	extra	extra		extra, extracapsular, extracellular, extracurricular, extradural
extra	ex	ex		extract, extractable, extractible, extraction, extractor
extro	extro	extro		extrospective, extroversion, extroversive, extrovert, extroverted
extr	exter	extr	extr	extreme, extremely, extremeness, extremism, extremist
ext	ex	ex	s	extant, extirpable, extirpate, extirpation
ext	ex	ex		extemporaneous, extemporaneously, extemporarily, extemporary, extempore
exu	ex	ex	s	exult, exultant, exultantly, exultation, exulting
exu	ex	ex		exurbia, exuberance, exuberant, exuberantly, exuberate
ex	ex	ex		exabit, exabyte, exbibit, exbibyte, exacerbate
e	ex	e		ebracteate, ebullient, ebullition, eburnation, eclair
grand	grand	grand		grandaunt, grandchild, granddad, granddaddy, granddaughter
gran	grand	gran		grandad
hyph	hypo	hyp		hyphema, hypha, hyphen, hyphenate,

Footprint	Prefix name	Character sequence to delete	Character sequence to insert	Sample instances
				hyphenation
hyp	hypo	hyp		hypaethral, hypanthium, hypesthesia, hypethral, hyponym
igni	igni	igni		ignitable, ignite, ignited, igniter, ignitible
ign	igni	ign		igneous, ignescent
ill	in	il		illume, illuminance, illuminant, illuminate, illuminated
imb	in	im		imbed, imbibe, imbiber, imbibing, imbibition
imm	in	im		immanence, immanency, immanent, immerse, immersion
imp	in	im		impact, impacted, impaction, impair, impaired
imp	en	im		improvable, improve, improved, improvement, improver
inan	inan	inan		inane, inanely, inanition, inanity
inb	in	in		inboard, inborn, inbound, inbred, inbreeding
industr	endo	indu		industrial, industrialisation, industrialise, industrialised, industrialism
infern	infern	infern		infernal, infernally, inferno
infer	infra	infer		inferior, inferiority, kine- prefix
infra	infra	infra		infra, infrahuman, inframaxillary, infrared, infrasonic
infra	in	in		infract, infraction, infrangible
initi	initi	initi		initial, initialisation, initialise, initialization, initialize
inl	in	in		inlaid, inland, inlay, inlet
inm	in	in		inmarriage, inmarry, inmate, inmost
inner	inner	inner		innermost, innersole
inn	in	in		innards, inner, inning, innings
inp	in	in		inpour, inpouring, input, inpatient
inq	in	in		inquest, inquietude, inquire, inquirer, inquiring
inr	in	in		inroad, inrush
insul	insul	insul		insulant, insular, insularism, insularity, insulate
integr	integr	integr		integer, integral, integrality, integrally, integrate
intellect	intellec	intellect		intellect, intellection, intellectual, intellectualisation, intellectualization
intellig	intellec	intellig		intelligence, intelligent, intelligently, intelligentsia, intelligibility
inter	inter	inter		inter, interact, interaction, interactional, interactive
inter	inter1	inter		interior, interiorise, interiorize, internal, internalisation
inte	in	in		integument, integumental, integumentary, intend, intended
intim	intim	intim		intima, intimacy, intimal, intimate, intimately
intra	intra	intra		intracapsular, intracellular, intracellular, intracerebral, intracranial
intro	intro	intro		intro, introduce, introduction, introductory, introit
inw	in	in		inward, inwardly, inwardness, inwards, inweave
in	in	in		inaugural, inaugurally, inaugurate, inauguration, incandesce
in	in	in		informatively, informatory, informed, informer, informercial
in	in	in		intoxicating, intoxication, intrench, intrenchment, intricacy
irr	in	ir		irradiate, irradiation, irregardless, irrigate
isol	insul	isol		isolate, isolation, isolator
kineto	kine	kineto		kinetochore, kinetosis
kinet	kine	kinet		kinetic
kine	kine	kine		kinematics, kinescope, kinesiology, kinesis

Footprint	Prefix name	Character sequence to delete	Character sequence to insert	Sample instances
kins	kin	kins		kinsfolk, kinsman, kinsperson, kinswoman
kin	kine	kin		kinaesthesia, kinaesthesis, kinaesthetic, kinanesthesia, kinesthesia
kin	kin	kin		kinfolk, kindred
metall	metal	metall		metallic, metallike, metallize, metalloid, metallurgic
metal	metal	metal		metal, metalhead, metalize, metalware, metalwork
meta	meta	meta		metabola, metabolic, metabolically, metabolise, metabolism
methyl	meth	methyl		methyl, methylated, methylbenzene, methyldopa, methylene
meth	meta	met		method, methodical, methodically, methodicalness, methodological
meth	meth	meth		methacholine, methacrylic, methamphetamine, methamphetamine, methane
metr	metr	metr		meter, metre, metric, metricate, metricise
met	meta	met		metempsychosis, metencephalon, metonym, metopion, metoprolol
misc	misc	misc		miscegenate, miscellanea, miscellany, miscible
miso	miso	miso		misogamy, misogynism, misogyny, misopedia
mis	miso	mis		misanthrope, misanthropy
mis	mis	mis		misaddress, misadventure, misadvise, misalign, misally
nonagen	nonagen	nonagen		nonagenarian
none	none	none		none, nonesuch, nonetheless, nonsuch
non	non	non		nones
obb	ob	obb		obbligato
obo	obo	obo		oboe, oboist
ob	ob	ob		obduracy, obdurate, obdurately, obedience, obedient
occ	ob	oc		occasion, occident, occipital, occiput, occlude
offic	op	of		office, officialdom, officialese, officiate, officious
off	off	off		offbeat, offhand, offhanded, offload, offprint
off	ob	of		offence, offend, offense, offensive, offer
opp	ob	op		opportune, opportunist, oppose, oppress, oppressor
ost	ob	os		ostensible, ostensive, ostensorium, ostentate, ostinato
ost	host	ost		ostler
para	para	para		parable, parabola, parabolic, parabolical, paraboloid
para	para1	para		parade, parader, paradiddle, parapet, parry
parent	par	parent		parent, parenteral
pari	par	pari		paries, parietal
pari	pari	pari		pari, parimutuel, parity, paripinnate
parl	parl	parl		parlance, parlay, parley, parliament, parlor
parol	parol	parol		parole, parolee
partheno	partheno	partheno		parthenocarpy, parthenogenesis, parthenogenetic, parthenogeny, parthenote
parti	parti	parti		parti, partial, partible, participant, participat
parturi	par	parturi		parturiency, parturient, parturition
parv	parv	parv		parve, parvis, parvo, parvo-virus
par	part	par		parboil, parcel, partake, parse, partner
par	para	par		paraesthesia, paraldehyde, paregmenon, paregoric, parenchyma
par	per	par		paramour, paramnesia, pardner, pardon, parfait
par	pari	par		par, parous
polar	pole	polar		polarimeter, polariscope, polarography

Footprint	Prefix name	Character sequence to delete	Character sequence to insert	Sample instances
polem	polem	polem		polemic, polemise, polemist, polemize, polemoniaceous
pole	pole	pole		poleax, poleaxe, polecat, pole, polestar
polic	poli	polic		police, policy
polit	poli	polit		politburo, polite, politic, polity, politesse
polen	pollen	polen		polenta, pollen
pollin	pollen	pollin		pollinate
pollu	pollu	pollu		pollute, pollution
polon	polon	polon		polonaise, polonium, polka
pol	pole	pol		polar, pollard
sub	sub	sub		subacid, subacute, subalpine, subaltern, subaquatic
succu	succ	succu		succulent
succ	sub	suc		succedaneum, succeed, success, successor, succinct
suff	sub	suf		suffer, suffice, sufficient, suffix, suffocate
sugg	sub	sug		suggest
summ	summ	summ		summate, summit
summ	sub	sum		summon, summons
supp	sub	sup		supplant, supple, supplejack, supplicate, supply
sust	sub	sus		sustain, sustenance, sustentacular, sustentation
syll	syn	syl		syllabary, syllabify, syllabise, syllable, syllabled
symb	syn	sym		symbiosis, symbiotic, symbol, symbolatry, symbology
symm	syn	sym		symmetry
symp	syn	sym		sympathectomy, sympathomimetic, sympathy, sympatry, sympetalous
syst	syn	sy		system, systematise, systole
unctu	unct	unctu		unctuous, unctuously, unctuousness
unct	unct	unct		unction
undula	undula	undula		undulant, undulate, undulation, undulatory
ungula	ungula	ungula		ungulate, ungulated, unguiculate, unguiculated, unguis
ungu	unct	ungu		unguent
uni	uni	uni		unicameral, unicellular, unicorn, unicuspid, unicycle
un	uni	un		unanimity, unanimous, unanimously, unary
un	un	un		until, unto

Appendix 50

Prefix translations

Regular prefixes

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
acantho	thorn	N.	flower	N.				
acet	vinegar	N.						
acro	sharp	ADJ.						
actino	ray	N.						
adeno	gland	N.						
aer	air	N.						
aero	air	N.						
algo	algebra	N.						
allo	other	ADJ.						
ambi	both	ADJ.						
amino	ammonia	N.						
amni	membrane	N.						
amphi	both	ADJ.						
amygdal	tonsil	N.						
angel	angel	N.						
angio	vessel	N.						
anthrop	human	N/A	man	N.				
anthropo	human	N/A	man	N.				
anim	live	ADJ.	life	N.				
apo	from	PREP.	away	ADV.				
aqua	water	N.						
arachno	spider	N.						
archae	old	ADJ.	ancient	ADJ.				
arche	old	ADJ.	ancient	ADJ.				
archi	chief	N/A	rule	V.				
arteri	artery	N.						
arterio	artery	N.						
arthro	hollow	ADJ.						
arti	skill	N.	art	N.	invention	N.		
astro	star	N.						
athero	porridge	N.						
audio	hear	V.						
augu	divination	N.						
auto	self	N.	automatic	ADJ.				
axi	axle	N.						
bacterio	bacteria	N.						
ball	throw	V.	ball	N.				
barb	beard	N.						
barbar	barbarian	N/A						
basidio	base	N.	bottom	N.				
basidio	base	N.						
bathy	deep	ADJ.						
bene	well	ADV.						
benzo	benzene	N.						
bi	twice	ADV.	two	ADJ.				

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
biblio	book	N.						
bio	life	N.						
blasto	sprout	N.						
bryo	moss	N.						
caco	bad	ADJ.						
cal	hot	ADJ.	heat	N.				
calci	lime	N.						
calli	beautiful	ADJ.	pretty	ADJ.				
calori	heat	N.						
cant	sing	V.						
carbo	coal	N.						
carcino	cancer	N.						
cardio	heart	N.						
carni	flesh	N.	meat	N.				
carpo	fruit	N.						
cata	down	A/P						
cent	hundred	ADJ.						
centr	centre	N.						
cephal	head	N.						
cephalo	head	N.						
chemo	chemistry	N.						
chlor	green	ADJ.	chlorine	N.				
chloro	green	ADJ.	chlorine	N.				
chole	bile	N.						
chor	choir	N.	land	N.				
chord	cord	N.						
chrom	colour	N.	chromium	N.				
chromat	colour	N.						
chromo	colour	N.						
chrono	time	N.						
chryso	gold	N/A						
circum	around	A/P						
claustro	shut	V.	close	V.	bolt	N.		
co	together	A/A						
coel	hollow	ADJ.						
cortico	bark	N.						
counter	against	PREP.						
cruci	cross	N.						
cryo	ice	N.	cold	ADJ.				
crypt	hidden	ADJ.	secret	ADJ.				
crypto	hidden	ADJ.	secret	ADJ.				
cteno	comb	N.						
culp	blame	V.						
cupro	copper	N.						
cur	care	N.						
cyano	blue	ADJ.	cyanide	N.				
cyber	virtual	ADJ.						
cycl	wheel	N.	circle	N.				
cyclo	wheel	N.	circle	N.				
cysto	bladder	N.						
cyto	cell	N.						

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
dacryo	tear	N.	weep	V.				
deca	ten	ADJ.						
deka	ten	ADJ.						
dermato	skin	N.						
dino	terrible	ADJ.						
diplo	double	ADJ.						
domi	house	N.	home	N.				
domin	lord	N.	master	N.				
dupl	double	ADJ.						
dyna	power	N.	force	N.				
dys	badly	ADV.	bad	ADJ.	ill	A/A		
ecto	outside	A/P	outer	ADJ.				
electr	electricity	N.						
electro	electric	ADJ.						
encephalo	brain	N.						
endo	inside	A/P	inner	ADJ.				
equi	equal	ADJ.						
ergo	work	N.						
erythro	red	ADJ.						
estro	frenzy	N.	impulse	N.				
extra	outside	A/P						
exuvia	undress	V.						
faeca	faeces	N.	stool	N.	shit	N.	feces	N.
fantas	imagination	N.	vision	N.				
febri	fever	N.						
feca	feces	N.	stool	N.	shit	N.	feces	N.
femto	quadrillionth	N.						
fibr	fibre	N.						
fibro	fibre	N.						
fiss	split	N/V						
flam	flame	N.						
fluoro	fluorine	N.						
foeto	embryo	N.	foetus	N.				
fond	melt	V.						
gall	cock	N.	French	ADJ.				
gam	marry	V.	mate	N/V				
gamet	mate	N/V	marry	V.	gamete	N.		
gastr	stomach	N.						
gastro	stomach	N.						
gen	heredity	N.	race	N.	kind	N.	sort	N.
gen	people	N.						
geo	earth	N.						
giga	billion	ADJ.	giant	ADJ.				
glycer	sweet	ADJ.						
glyco	sweet	ADJ.						
granul	grain	N.						
grapho	write	V.	draw	V.				
guaran	guarantee	N/V						
gymn	bare	ADJ.	naked	ADJ.				
gyn	woman	N.						
haem	blood	N.						

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
haemato	blood	N.						
haemo	blood	N.						
halo	salt	N.						
hecto	hundred	ADJ.						
helio	sun	N.						
hem	blood	N.						
hemat	blood	N.						
hemato	blood	N.						
hemi	half	ADJ.						
hemo	blood	N.						
hepato	liver	N.						
hetero	other	ADJ.						
hexa	six	ADJ.						
hind	back	N.						
hist	tissue	N.						
holo	whole	ADJ.						
homeo	same	ADJ.						
homo	same	ADJ.						
horo	hour	N.						
hydr	water	N.	hydrogen	N.				
hydro	water	N.	hydrogen	N.				
hygro	wet	ADJ.	moist	ADJ.				
hyper	above	A/P	over	A/P				
hypno	sleep	N/V						
hypo	under	A/P	beneath	A/P				
icono	picture	N.						
ideo	idea	N.						
idio	private	ADJ.	personal	ADJ.				
immuno	immune	ADJ.						
inter	among	PREP.	between	A/P				
intra	inside	A/P						
iodo	purple	ADJ.	iodine	N.				
iso	equal	ADJ.						
kara	empty	ADJ.						
karyo	kernel	N.						
kerat	hair	N.						
kerato	hair	N.						
keto	acetone	N.						
kilo	thousand	ADJ.						
lact	milk	N.						
laryngo	larynx	N.						
legi	law	N.	read	V.				
lent	slow	ADJ.						
lenti	lentil	N.	lens	N.				
lepido	scale	N.						
lepto	small	ADJ.	little	ADJ.				
leuco	white	ADJ.						
leuko	white	ADJ.						
lipo	fat	ADJ.						
litho	stone	N.	rock	N.				
loco	place	N.						

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
logo	word	N.	idea	N.				
loxo	oblique	ADJ.						
lyc	wolf	N.						
lymph	lymph	N.						
lympho	lymph	N.						
lyso	loose	ADJ.						
macro	long	ADJ.						
magni	big	ADJ.	large	ADJ.	great	ADJ.		
magneto	magnet	N.						
mal	bad	ADJ.	badly	ADV.				
man	hand	N.						
matri	mother	N.						
med	middle	N.						
mega	big	ADJ.	million	ADJ.	large	ADJ.		
megalo	big	ADJ.	large	ADJ.				
melan	black	ADJ.						
meri	part	N.						
mero	part	N.						
meso	middle	N.	medium	ADJ.				
micr	little	ADJ.	small	ADJ.				
micro	little	ADJ.	small	ADJ.				
mid	middle	N.						
milli	thousand	ADJ.						
mini	little	ADJ.	small	ADJ.				
moll	soft	ADJ.						
mon	single	ADJ.	alone	ADJ.	only	ADJ.		
mono	single	ADJ.	alone	ADJ.	only	ADJ.		
mont	mountain	N.	hill	N.				
mort	death	N.						
muco	snot	N.						
multi	many	ADJ.						
muta	change	V.						
myco	fungus	N.						
myel	marrow	N.						
myelo	marrow	N.						
myo	muscle	N.	mouse	N.	shut	ADJ.		
myria	ten thousand	ADJ.	many	ADJ.				
myric	tamarisk	N.						
nano	dwarf	N.	tiny	ADJ.	microscopic	ADJ.		
neo	new	ADJ.	young	ADJ.				
nebul	cloud	N.	mist	N.				
necro	corpse	N.						
neg	deny	V.	not	ADV.				
nephro	kidney	N.						
neur	nerve	N.						
neuro	nerve	N.						
nitr	nitrogen	N.						
nitro	nitrogen	N.						
nomo	law	N.	coin	N.				
nucle	nucleus	N.						
nucleo	nucleus	N.						

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
nud	naked	ADJ.						
nympho	bride	N.	sex	N.	nymph	N.		
oct	eight	ADJ.						
oestro	frenzy	N.	impulse	N.				
olig	few	ADJ.						
omni	all	ADJ.	every	ADJ.				
ora	beg	V.	pray	V.				
orchi	testicle	N.						
ortho	true	ADJ.	right	ADJ.				
oscillo	swing	V.						
osteo	bone	N.						
ox	sharp	ADJ.	bitter	ADJ.	oxygen	N.		
oxy	sharp	ADJ.	bitter	ADJ.	oxygen	N.		
pachy	thick	ADJ.						
palaeo	old	ADJ.	ancient	ADJ.				
paleo	old	ADJ.						
palin	again	ADV.						
pan	all	ADJ.	every	ADJ.	Pan	N.		
patho	suffer	V.	experience	N.				
patri	father	N.						
pen	almost	ADV.						
ped	child	N.						
pedi	foot	N.						
pent	five	ADJ.						
penta	five	ADJ.						
per	through	A/P	thorough	ADJ.				
peri	about	A/P	around	A/P				
petro	rock	N.	stone	N.				
phanero	appear	V.						
pharmac	drug	N.	poison	N.				
pheno	phenol	N.	shining	ADJ.				
phenyl	phenol	N.	shining	ADJ.				
phil	love	V.						
phon	voice	N.						
phosph	phosphorus	N.						
photo	light	N.	photography	N.				
phyto	plant	N.						
pico	trillionth	N.						
pinnat	winged	ADJ.	feathered	ADJ.				
pinni	fin	N.						
plan	flat	ADJ.						
planti	plant	N.	sole	N.				
plas	mold	N.						
pleon	more	A/A	enough	A/A				
plu	more	A/A	most	ADJ.	many	ADJ.	much	A/A
pneumo	lung	N.	breath	N.	air	N.	wind	N.
pogoni	beard	N.						
poly	many	ADJ.						
popul	people	N.						
porphyri	purple	ADJ.	porphyry	N.				
port	carry	V.	gate	N.	port	N.	bring	V.

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
post	putrid	ADJ.	positive	ADJ.				
post	after	A/P						
pre	before	A/P						
pro	for	PREP.	before	A/P				
prote	protein	N.						
proto	first	ADJ.						
pseudo	false	ADJ.						
psych	mind	N.						
psycho	mind	N.						
ptero	wing	N.						
pterido	wing	N.						
pur	for	PREP.						
puta	think	V.						
putr	rot	V.						
pyro	fire	N.						
quadr	four	ADJ.						
quart	fourth	ADJ.						
quater	four	ADJ.						
radio	radiation	N.	radio	N.	ray	N.		
re	back	ADV.	again	ADV.				
reg	rule	V.						
reti	net	N.						
retro	backwards	ADV.	back	ADV.				
rhabdo	stick	N.						
rhin	nose	N.						
rhino	nose	N.						
rhizo	root	N.						
sacr	sacred	ADJ.						
sal	salt	N.						
sapro	putrid	ADJ.						
sarco	flesh	N.						
satis	enough	A/A						
scal	scale	N.	ladder	N.				
scler	hard	ADJ.						
sclero	hard	ADJ.						
se	apart	A/A	separate	ADJ.	without	PREP.		
secret	hidden	ADJ.						
sei	shake	V.						
semi	half	ADJ.						
sen	sense	V.	feel	V.				
sequ	follow	V.						
sider	star	N.						
silic	silicon	N.	flint	N.				
simpl	simple	N.	single	ADJ.				
sinistr	left	N.						
somato	body	N.						
son	sound	N.						
spectro	spectrum	N.						
sperm	seed	N.						
spermat	seed	N.						
spher	ball	N.	round	ADJ.	globe	N.		

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
spir	breathe	V.	coil	N/V				
spongi	sponge	N.						
spor	spore	N.						
statu	stand	N.						
statu	set up	V.						
sterco	dung	N.						
stom	mouth	N.						
stomat	mouth	N.						
strepto	twisted	ADJ.						
strob	whirl	V.						
styr	resin	N.						
sulf	sulfur	N.	sulphur	N.				
sulph	sulphur	N.	sulfur	N.				
super	above	A/P	on	A/P	over	A/P		
supra	above	A/P	on	A/P	over	A/P		
sur	on	A/P	above	A/P	over	A/P		
swa	self	N.						
syrin	pipe	N.						
syn	with	PREP.						
tach	fast	ADJ.						
techn	skill	N.	invention	N.				
tele	far	A/A						
teleo	end	N.						
telo	end	N.						
temp	time	N.	weather	N.				
terato	marvel	N.						
tetr	four	ADJ.						
tetra	four	ADJ.						
ther	beast	N.	animal	N.	fierce	ADJ.	wild	ADJ.
therm	heat	N.						
thermo	heat	N.						
thromb	clot	V.						
thrombo	clot	V.						
thyro	thyroid	N.						
trans	across	A/P						
tri	three	ADJ.						
trop	turn	V.						
turb	turmoil	N.	crowd	N.				
tyrann	tyrant	N.	king	N.				
ultim	last	A/A						
ultra	beyond	A/P						
under	under	A/P	beneath	A/P				
ur	urine	N.	piss	V.				
vapor	steam	N.						
vaso	vessel	N.						
ver	real	ADJ.	TRUE	ADJ.				
vern	spring	N.						
verb	word	N.						
verd	green	ADJ.						
vermi	worm	N.						
vibra	shake	V.	vibrate	V.				

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
vill	house	N.	village	N.	town	N.		
vol	want	V.	wish	V.				
volcan	volcano	N.						
with	with	PREP.						
xeno	strange	ADJ.						
xero	dry	ADJ.						
zoo	animal	N.						
zygo	yoke	N.						
zymo	leaven	N.	yeast	N.				

Irregular prefixes

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
a								
a1	without	PREP.						
ab	from	PREP.	away	ADV.				
abba	father	N.						
acri	sharp	ADJ.						
ad	to	PREP.	at	PREP.				
adolesc	teen	N/A						
adult	adult	N/A						
ali	other	ADJ.						
all	all	ADJ.						
allo	other	ADJ.						
alt	high	ADJ.						
altr	other	ADJ.						
am	love	N/V	like	V.				
amp	amp	N.						
ana	up	A/P	back	ADV.	against	PREP.	again	ADV.
ana	to	PREP.	through	A/P				
andro	man	N.	male	N/A				
anemo	wind	N.						
ank	narrow	ADJ.						
ann	year	N.						
annu	ring	N.						
ano	anus	N.						
ante	before	A/P						
antho	flower	N.						
anti	against	PREP.						
antiqu	old	ADJ.						
aperi	open	V.						
aphro	sex	N.						
api	bee	N.						
apo	from	PREP.	away	ADV.				
apt	apt	ADJ.						
arc	inverse	ADJ.						
arch	chief	N/A						
be								
cata	down	A/P	against	PREP.	wrongly	ADV.		
col	glue	N.						
coll	neck	N.						

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
coll1	cabbage	N.						
coll2	coal	N.						
coll3	colic	N.						
com	revel	V.						
comme	as	PREP.	how	ADV.				
con	with	PREP.	together	ADV.				
cone	cone	N.						
contra	against	PREP.						
cop	cut	V.						
corr	run	V.						
de	from	PREP.	away	ADV.	down	A/P	about	A/P
de	off	A/P	among	PREP.	completely	ADV.		
deb	owe	V.						
deco	nice	ADJ.						
dec	ten	ADJ.						
deed	done	V/A						
dei	god	N.	God	N.				
deka	ten	ADJ.						
del	destroy	V.						
dem	people	N.						
demi	half	ADJ.						
demon	spirit	N.						
dendr	tree	N.						
denti	tooth	N.						
derm	skin	N.						
despot	lord	N.						
deterior	worse	A/A						
deu	two	ADJ.						
deuter	second	ADJ.						
dextro	right	N.						
di	twice	ADV.						
di1	day	N.						
dia	across	A/P	through	A/P	thorough	ADJ.		
die	die	V.						
dis	from	PREP.	away	ADV.	down	A/P	about	A/P
dis	off	A/P	among	PREP.	completely	ADV.		
disco	plate	N.						
ec	out	ADV.	out of	PREP.				
ecclesi	church	N.						
echino	spiny	ADJ.						
echo	echo	N.						
eco	live	V.						
ecto	outside	A/P	outer	ADJ.				
en	in	A/P	into	PREP.				
en								
endo	inside	A/P	inner	ADJ.				
eno	one	ADJ.						
enter	inside	A/P	among	PREP.	between	A/P		
entero	gut	N.	intestine	N.				
ento	inside	A/P						
entomo	insect	N.						

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
epan	again	ADV.						
epaul	shoulder	N.						
epi	on	A/P						
err	wander	V.						
eu	well	ADV.						
ex	out	ADV.	out of	PREP.				
exo	outside	A/P						
exter	outside	A/P						
extra	outside	A/P						
extro	outward	A/A						
grand								
host	host	N.						
hyper	above	A/P	over	A/P				
hypo	under	A/P	beneath	A/P				
igni	fire	N.						
in	in	A/P	into	PREP.				
inan	empty	ADJ.						
infern	below	ADV.						
infra	within	A/P						
initi	begin	V.	start	N/V				
inner	inner	ADJ.						
insul	island	N.						
integr	whole	ADJ.						
intellec	intelligent	ADJ.						
inter	among	PREP.	between	A/P				
inter1	inside	A/P						
intim	intimate	ADJ.						
intra	inside	A/P						
intro	inward	A/A						
kin	family	N.						
kine	movement	N.						
meta	after	A/P	beyond	A/P	changed	ADJ.		
metal	metal	N/A						
meth	methyl	N.						
metr	measure	N/V						
mis	badly	ADV.	wrong	A/A				
misc	mix	N/V						
miso	hate	N/V						
non	ninth	ADJ.						
nonagen	ninety	ADJ.						
none	none	N.						
ob	in front of	PREP.	against	PREP.	towards	PREP.	before	A/P
ob	about	A/P						
obo	oboe	N.						
off	off	A/P						
op	work	N.						
par	birth	N.						
para	alongside	A/P	beyond	A/P	changed	ADJ.	contrary	ADJ.
para	beside	PREP.	near	A/P				
para1	prepare	V.						
pari	equal	ADJ.						

Prefix	Translation	POS	Translation	POS	Translation	POS	Translation	POS
parl	talk	V.						
parol	word	N.						
part	part	N.						
partheno	virgin	N.						
parti	part	N.						
parv	little	ADJ.	small	ADJ.				
per	through	A/P	thorough	ADJ.				
pole	pole	N.						
polem	war	N.						
poli	state	N.	city	N.				
pollen	flour	N.	pollen	N.				
pollu	pollution	N.						
polon	Polish	ADJ.						
sub	under	A/P	beneath	A/P				
succ	juice	N.						
summ	total	N/A						
syn	with	PREP.						
un	not	ADV.						
unct	anoint	V.						
under	under	A/P	beneath	A/P				
undula	wave	N.						
ungula	hoof	N.	nail	N.				
uni	single	ADJ.	one	ADJ.				

Appendix 51

1st. secondary prefix set as ordered by the optimal heuristic

over	re	out	under	micro
counter	super	back	semi	pro
fore	s	poly	hyper	down
cross	pre	neuro	trans	auto
post	multi	side	radio	photo
cyto	for	qu	tri	after
electro	mega	mono	c	thermo
endo	hydro	pseudo	tele	osteo
paleo	co	milli	lxx	squ
per	p	iso	psycho	angio
hetero	cyber	syn	circum	ma
ca	tetra	aero	palaeo	bi
macro	adeno	qua	pyro	nephro
jack	car	nitro	ba	blasto
lymph	b	t	la	ultra
kilo	st	xeno	sarco	acro
sun	tran	ga	cata	kerato
immuno	matri	mo	phyto	homo
equi	peri	gra	myco	amphi
hemato	proto	arthro	do	patri
mon	apo	necro	biblio	strepto
diplo	karyo	ch	up	cardio
ortho	pla	hydr	li	ne
actino	ha	pe	radi	ergo
chole	phenyl	ver	vi	whi
war	fo	chemo	hecto	bur
zoo	mini	helio	tr	cyclo
dys	megalo	wa	acet	ra
plough	zymo	cha	ja	crypto
thyro	with	lo	hypno	retr
gr	sp	sc	hind	haemo
rhizo	quater	rhabdo	carcino	zygo
terato	volcan	th	hypo	pa
se	hydroxy	he	bo	haemato
ho	lipo	fibro	va	lxxx
thrombo	homeo	in	pr	sa
swa	hemat	fluoro	xx	me
bomb	ove	retro	fla	myo
laryngo	bio	ta	spectro	synchro
xxx	astro	no	bar	m
na	tur	squa	le	oxy
aqua	erythro	lenti	requi	hepato
tra	da	te	pneumo	moor
sea	fl	tetr	corn	penta
socio	bladder	fibrino	di	dra
man	br	g	bra	rein

ski	sur	pan	sh	mid
myel	lepto	lepido	sequ	idio
omni	secre	seve	acantho	icono
litera	papill	amni	lexico	modul
pancrea	popul	albin	foeto	sapro
athero	butter	cytoplas	gonadotrop	guaran
lepidopter	nerit	phantas	protozo	underli
valvul	bathyscap	cockle	dacryo	exuvia
gliste	hove	iconoclas	mollus	overhea
panthe	taff	ve	al	a
po	litho	cla	f	nucleo
ka	to	gastr	ar	pur
mi	chrom	fur	bla	pen
gastro	qui	myelo	pal	anthropo
nano	sca	thro	neur	muco
count	pass	micr	vermi	oto
bacterio	oct	sta	palae	hemo
wood	domi	arterio	chromo	phospho
therm	hist	myxo	aer	vaso
chlo	chi	audi	xero	benefi
dyna	water	red	sal	iodo
colum	hum	lent	hexa	nebul
rever	fantas	cent	eth	upst
amino	silic	l	ste	cro
chloro	un	cortico	basidio	bocc
breech	ginger	jell	malle	meteor
signor	lympho	fa	mar	fil
ki	sla	ro	encephalo	vill
audio	techno	vol	gro	the
port	pent	meso	benzo	drago
eel	patho	vibra	cur	cr
bill	procto	simpl	beig	briar
cedar	chilias	curle	oscillo	pogoni
porphyri	shallo	thimble	through	phono
cryo	cros	orchi	har	sno
nympho	ornitho	trave	there	asco
wi	rhin	top	gar	chryso
cyano	domin	cor	ya	calli
temp	ye	blin	rhino	lin
cre	so	fe	cal	kha
electr	psych	quadr	immun	thromb
cephal	anthrop	acanth	arteri	vul
nucle	scler	glycer	umb	cruci
pharmac	sulph	amygdal	calori	ethan
granul	xantho	chris	femto	maxill
phyco	sigmoi	suprem	vesic	allo
gyro	petro	scen	trache	acryl
angeli	bacchan	bicolo	botuli	derri
heredit	ichthyo	igno	monochrom	ocul
oneir	orbi	porphyr	radiotele	seren

synthe	academ	acous	aesthe	amphibol
aneur	angiocar	argenti	baptis	batholit
benedic	binuclea	bronchiol	campanul	cannul
cataplas	catapul	centesi	cervi	chlorophy

Appendix 52

Linking vowel exceptions and reverse linking vowel exceptions

Linking vowel exceptions

Prefix with a linking vowel	Stem with a missing initial vowel	Prefix with a linking vowel	Stem with a missing initial vowel	Prefix with a linking vowel	Stem with a missing initial vowel	Prefix with a linking vowel	Stem with a missing initial vowel
hetero	ecious	trans[21]	cend	cephalo	ridine	audi	ble
hetero	icous	trans	cendental	cephalo	thin	audi	le
hetero	sis	trans	cribe	leuko	ma	febri	le
hydro	id	trans	cript	andro	ena		
hydro	ps	trans	criptase	andro	enid		
hydro	xide	trans	ect	andro	ecium		
hydro	xy	trans	ept	dextro	rsal		
hydro	xybenzene	trans	exual	dextro	rse		
hydro	xychloroquine	trans	om	dextro	se		
hydro	xyl	trans	onic	dextro	us		
hydro	xymethyl	trans	pire	dis	hevel		
hydro	xyproline	trans	ubstantiate	entero	ptosis		
hydro	xytetracycline	zoo	psia	parti	cle		
hydro	xyzine	apo	dous	carcino	ma		
iso	smotic	athero	ma	carcino	matous		
micro	glia	chryso	pid	litho	ps		
micro	gliacyte	crypto	rchidism	mono	cle		
neuro	glia	crypto	rchidy	mono	cled		
neuro	gliacyte	crypto	rchism	mono	dy		
neuro	ma	hemo	ptysis	mono	ecious		
neuro	matous	hepato	ma	mono	estrous		
osteo	ma	hexa	ne	mono	icous		
co	ver	hexa	ngular	mono	rchidism		
co	vert	iodo	psin	mono	rchism		
ergo	dic	myo	ma	mono	vular		
haemo	ptysis	myo	pe	mono	xide		
helio	psis	myo	pia	myelo	ma		
macro	glia	necro	psy	nano	phthalmos		
ortho	ptic	penta	cle	orchi	tis		
ortho	ptist	penta	ngle	petro	latum		
paleo	ntology	penta	thlete	petro	leum		
peri	sh	penta	thlon	radio	pacity		
pre	dnisolone	quater	nion	radio	paque		
pre	dnisone	quater	nity	amphi	sbaena		
psycho	did	xero	ma	blasto	ma		
sarco	ma	zygo	ma	ambi	ent		
sarco	ptid	astro	glia	holo	nym		
tele	ncephalon	carbo	xyl	palae	stra		
tele	vangelism	carbo	xylic	palae	tiology		

[21] The same principle applies even though 's' is not a vowel.

Reverse linking vowel exceptions

Prefix without linking vowel	Stem with erroneous initial vowel	Prefix without linking vowel	Stem with erroneous initial vowel	Prefix without linking vowel	Stem with erroneous initial vowel
lymph	oblast	mon	olith	chlor	oacetophenone
lymph	ocyte	mon	olithic	chlor	obenzene
lymph	ocytopenia	mon	ologist	chlor	obenzylidenemalononitrile
lymph	ocytosis	mon	ologue	chlor	ofluorocarbon
lymph	ogranuloma	mon	omania	chlor	oform
lymph	ography	mon	omaniac	chlor	ofucin
lymph	oid	mon	omaniacal	chlor	ophyl
lymph	okine	mon	omer	chlor	ophyll
lymph	oma	mon	ometallic	chlor	ophyllose
lymph	openia	mon	omorphemic	chlor	ophyte
lymph	opoiesis	mon	oneuropathy	chlor	opicrin
mon	oamine	mon	onuclear	chlor	oplast
mon	oatomic	mon	onucleate	chlor	oprene
mon	oblast	mon	onucleosis	chlor	oquine
mon	ocarboxylic	mon	ophony	chlor	osis
mon	ocarp	mon	oplane	chlor	othiazide
mon	ocarpic	mon	oplegia	chlor	otic
mon	ochromasy	mon	oploid	chrom	oblastomycosis
mon	ochromat	mon	opoly	chrom	ogen
mon	ochrome	mon	opsony	chrom	olithography
mon	ochromia	mon	opteral	chrom	ophore
mon	ocline	mon	orail	chrom	oplast
mon	oclonal	mon	orchidism	chrom	osomal
mon	ocot	mon	orchism	chrom	osome
mon	ocotyledon	mon	osaccharide	chrom	osphere
mon	ocracy	mon	osaccharose	domin	ie
mon	oculture	mon	osemy	domin	ion
mon	ocycle	mon	osomy	haem	atal
mon	ocyte	mon	osyllabic	haem	atemesis
mon	ocytosis	mon	osyllable	haem	atinic
mon	oecious	mon	otheism	haem	atite
mon	oestrous	mon	otone	haem	aturia
mon	ogamy	mon	otreme	man	ual
mon	ogenesis	mon	otype	man	ufactory
mon	ogenic	mon	ounsaturated	man	ufacture
mon	ogram	mon	ovalent	man	ul
mon	ograph	mon	ovular	man	umit
mon	ogyny	mon	ozygotic	man	ure
mon	ohybrid	acet	one	man	us
mon	ohydrate	acet	onemia	man	uscript
mon	oicous	acet	onuria	pen	eplain
mon	olatry	acet	ophenetidin	pen	eplane
mon	olingual	acet	ose	pent	obarbital

Prefix without linking vowel	Stem with erroneous initial vowel	Prefix without linking vowel	Stem with erroneous initial vowel	Prefix without linking vowel	Stem with erroneous initial vowel
pent	ode	chromat	ogram	part	ttime
pent	ose	chromat	ography	part	ty
psych	edelia	dyna	mise	pole	lard
quadr	ant	dyna	mite	part	tsong
quadr	aphony	fibr	eboard	amni	ote
quadr	asonic	fibr	eglass	amygdal	oid
quadr	ate	fibr	eoptic	amygdal	otomy
quadr	ature	hist	ocompatibility	archae	obacteria
quadr	iceps	hist	ogram	archae	ology
quadr	ilateral	hist	oincompatibility	archae	opteryx
quadr	ipara	hist	ology	archae	ornis
quadr	ipartite	hist	one	archae	ozoic
quadr	iphonic	oct	agon	gen	ocide
quadr	iplegia	oct	ahedron	gen	oise
quadr	iplegic	oct	al	gen	omics
quadr	isonic	oct	ameter	gen	otype
quadr	ivium	oct	ane	gen	tamicin
quadr	uped	oct	angular	gen	teel
quadr	uple	oct	ave	gen	tile
quadr	uplet	oct	avo	gen	tle
quadr	uplex	oct	ogenarian	gen	tly
quadr	uplicate	oct	onary	gen	trify
quadr	upling	oct	opod	glycer	ogel
rhizo	ctinia	oct	opus	glycer	ogelatin
sal	icylate	oct	oroon	granul	ocyte
scler	edema	oct	osyllabic	granul	ocytopenia
scler	oderma	oct	osyllable	keto	nemia
scler	ometer	oct	uple	keto	nuria
scler	oprotein	silic	ide	orchi	dectomy
scler	osed	silic	ious	orchi	opexy
scler	osis	demon	olatry	pharmac	ogenetics
scler	otic	dendr	obium	pharmac	okinetics
scler	otinia	disco	ography	pharmac	ology
scler	otium	disco	oid	pharmac	opeia
scler	otomy	disco	oidal	pharmac	opoeia
simpl	eton	disco	omycete	ver	isimilar
pneumo	nectomy	disco	otheque	ver	isimilitude
pneumo	nia	ecclesi	astic	ver	itable
pneumo	nitis	ecclesi	ology	ver	ity
pneumo	noconiosis	epan	alepsis	arche	opteryx
carbo	naceous	epan	aphora	olig	ochaete
carbo	nado	ex	otic	olig	oclase
carbo	nara	ex	otism	olig	odendrocyte
carbo	nate	in	nards	spher	ocyte
carbo	nyl	in	ning	spir	ochaete
carbo	nylic	part	tner	spir	ochete

Prefix without linking vowel	Stem with erroneous initial vowel	Prefix without linking vowel	Stem with erroneous initial vowel	Prefix without linking vowel	Stem with erroneous initial vowel
spir	ogram	bath	olith	melan	oblast
spir	ograph	bath	yscape	melan	ocyte
ther	opod	bath	yscaph	phil	ologue
ur	obilinogen	bath	yscaphe	phil	omath
ur	ochord	bath	ysphere	phon	ogram
ur	okinase	centr	ifuge	phon	ograph
ur	olith	centr	omere	prote	osome
din	osaur	centr	osome	tach	ogram
hal	ophyte	coel	iac	tach	ograph
spor	ocarp	coel	ostat	techn	ocrat
spor	ophore	cycl	amen	techn	ophobe
spor	ophyl	cycl	es/second	trop	onym
spor	ophyll	graph	ospasm	trop	opause
spor	ophyte	gymn	osperm	trop	osphere
aqu	ilege	gyn	obase	chor	eograph
arch	itect	gyn	ophore	pinn	iped
arch	itrave	lact	ifuge		
arch	osaur	lact	ogen		

Appendix 53

Secondary suffix stripping stoplist

Original word	Original POS	De-suffixed word	De-suffixed POS	Original word	Original POS	De-suffixed word	De-suffixed POS
aspirate	VERB	aspire	VERB	pappa	NOUN	pappus	NOUN
castrate	VERB	caster	NOUN	tala	NOUN	talus	NOUN
nominative	ADJECTIVE	nominate	VERB	tantra	NOUN	tantrum	NOUN
truant	ADJECTIVE	true	VERB	vara	NOUN	varus	NOUN
pa	NOUN	pus	NOUN	villa	NOUN	villus	NOUN
placoid	ADJECTIVE	place	NOUN	petition	NOUN	pet	VERB
tineoid	NOUN	tine	NOUN	acid	NOUN	ace	ADJECTIVE
aroid	NOUN	are	NOUN	fell	NOUN	fall	VERB
aroid	ADJECTIVE	are	NOUN	fell	ADJECTIVE	fall	VERB
choroid	NOUN	chore	NOUN	pall	VERB	pal	VERB
mastoid	NOUN	mast	NOUN	sold	ADJECTIVE	sell	VERB
mastoid	ADJECTIVE	mast	NOUN	solid	NOUN	sole	ADJECTIVE
archil	NOUN	arch	NOUN	sparid	NOUN	spare	ADJECTIVE
stridor	NOUN	stride	VERB	sultana	NOUN	sultan	NOUN
tailor	NOUN	tail	VERB	billyo	NOUN	billy	NOUN
pallor	NOUN	pal	VERB	bracero	NOUN	bracer	NOUN
signor	NOUN	sign	VERB	dinero	NOUN	diner	NOUN
minor	NOUN	mine	VERB	folio	NOUN	folie	NOUN
honor	NOUN	hone	VERB	lazaretto	NOUN	lazaret	NOUN
door	NOUN	do	VERB	magneto	NOUN	magnet	NOUN
censor	NOUN	cense	VERB	medico	NOUN	medic	NOUN
cursor	NOUN	curse	VERB	morello	NOUN	morel	NOUN
savor	NOUN	save	VERB				
salvor	NOUN	salve	VERB				
saw	NOUN	see	VERB				
pallor	NOUN	pall	VERB				
abaca	NOUN	abacus	NOUN				
actinia	NOUN	actinium	NOUN				
ala	NOUN	alum	NOUN				
ana	NOUN	anus	NOUN				
anna	NOUN	annum	NOUN				
asteroid	NOUN	aster	NOUN				
asteroid	ADJECTIVE	aster	NOUN				
basilar	ADJECTIVE	basil	NOUN				
bola	NOUN	bolus	NOUN				
calla	NOUN	callus	NOUN				
chiasma	NOUN	chiasmus	NOUN				
dura	NOUN	durum	NOUN				
lota	NOUN	lotus	NOUN				
mara	NOUN	marum	NOUN				
mina	NOUN	minus	NOUN				
pallor	NOUN	pal	VERB				

Appendix 54

Final suffixation reprieves

Stem	POS	Suffix 1	Suffix 2	Suffix 3
plane	NOUN	et	ar	ula
arm	NOUN	et	illa	
bulb	NOUN	ar	il	
face	NOUN	et	ula	
fuse	NOUN	iform	il	
gob	NOUN	et	let	
medic	NOUN	ate	o	
out	NOUN	let	ward	
prime	NOUN	ula	o	
scale	NOUN	ar	ar	
terce	NOUN	et	el	
turbine	NOUN	ate	ate	
yob	NOUN	o	o	
acerb	ADJECTIVE	ate		
acne	NOUN	iform		
alien	VERB	ee		
amble	NOUN	ulate		
annexa	NOUN	al		
arcane	ADJECTIVE	um		
argent	NOUN	ite		
argil	NOUN	ite		
baa	NOUN	s		
bar	VERB	ator		
barb	NOUN	el		
bard	NOUN	ic		
barkeep	NOUN	er		
basin	NOUN	et		
bean	NOUN	o		
bedsit	NOUN	er		
beth	NOUN	el		
billy	NOUN	o		
blank	NOUN	et		
blanket	VERB	t		
boneset	NOUN	er		
bookmark	NOUN	er		
bowl	NOUN	s		
bract	NOUN	let		
brave	NOUN	o		
breve	NOUN	et		
brief	NOUN	s		
bursa	NOUN	itis		
cabin	NOUN	et		
cane	NOUN	ella		
cant	NOUN	o		
car	NOUN	ry		
cardsharp	NOUN	er		
chiasmus	NOUN	a		

Stem	POS	Suffix 1	Suffix 2	Suffix 3
chick	NOUN	en		
chimneysweep	NOUN	er		
chrism	NOUN	ist		
christ	NOUN	ella		
copal	NOUN	ite		
crate	VERB	ate		
cube	NOUN	iform		
custody	NOUN	ian		
cyst	NOUN	itis		
date	VERB	ate		
dick	NOUN	y		
dig	NOUN	s		
dock	NOUN	et		
dote	VERB	age		
doublet	NOUN	on		
down	NOUN	ward		
dragon	NOUN	et		
drib	NOUN	let		
drug	NOUN	et		
drupe	NOUN	let		
dura	NOUN	ral		
durum	NOUN	a		
dyad	NOUN	ic		
ebon	ADJECTIVE	y		
empire	NOUN	ic		
ester	NOUN	one		
event	NOUN	ual		
fabric	NOUN	ate		
falanga	NOUN	ist		
faun	NOUN	na		
feist	NOUN	y		
fenestra	NOUN	ral		
flint	ADJECTIVE	nt		
flue	NOUN	id		
formic	ADJECTIVE	ate		
frequent	VERB	t		
front	NOUN	let		
galax	NOUN	ctic		
gate	VERB	ate		
gerbil	NOUN	le		
gingiva	NOUN	itis		
globe	NOUN	al		
gorge	NOUN	et		
graph	NOUN	ology		
grate	VERB	ate		
gun	NOUN	el		
gyre	NOUN	o		
habit	NOUN	us		
haem	NOUN	ic		
hate	VERB	ate		
herb	NOUN	al		

Stem	POS	Suffix 1	Suffix 2	Suffix 3
host	NOUN	el		
iridesce	VERB	scent		
iron	ADJECTIVE	y		
joint	ADJECTIVE	nt		
junk	NOUN	et		
lap	NOUN	et		
lave	VERB	ation		
lee	NOUN	s		
lie	NOUN	ar		
line	NOUN	ear		
lingua	NOUN	ist		
lively	ADJECTIVE	hood		
lobe	NOUN	ar		
lock	NOUN	et		
lure	NOUN	id		
luster	NOUN	ate		
magnet	NOUN	o		
maid	NOUN	en		
marine	NOUN	er		
mastic	NOUN	ate		
mean	NOUN	s		
meme	NOUN	o		
meteor	NOUN	ology		
millenary	NOUN	ian		
miller	NOUN	ite		
mime	NOUN	o		
mint	ADJECTIVE	nt		
miser	NOUN	ery		
mix	NOUN	ology		
mod	NOUN	s		
myth	NOUN	ic		
native	ADJECTIVE	ity		
neck	NOUN	let		
nine	ADJECTIVE	ety		
note	VERB	tion		
nub	NOUN	y		
numeric	ADJECTIVE	ous		
nymph	NOUN	o		
ohm	NOUN	ic		
old	NOUN	en		
organ	NOUN	ise		
palm	NOUN	ar		
pater	NOUN	ology		
peck	NOUN	ish		
pen	VERB			
phyllo	NOUN	de		
pink	NOUN	o		
pious	ADJECTIVE	ity		
pique	VERB	uant		
plate	VERB	ate		
pop	NOUN	et		

Stem	POS	Suffix 1	Suffix 2	Suffix 3
porn	NOUN	o		
prick	NOUN	et		
prune	NOUN	o		
pseud	NOUN	o		
pupil	NOUN	ary		
quantal	ADJECTIVE	ity		
ramp	VERB	ant		
ratch	NOUN	et		
rhythm	NOUN	ic		
rich	NOUN	s		
ropewalk	NOUN	er		
rose	NOUN	illa		
round	NOUN	el		
ruth	NOUN	ful		
sabot	NOUN	age		
salve	VERB	or		
saury	NOUN	ian		
seism	NOUN	ic		
seven	ADJECTIVE	ty		
sext	NOUN	et		
short	NOUN	s		
soph	NOUN	ism		
sot	NOUN	ish		
statue	NOUN	ary		
tart	NOUN	let		
ten	NOUN	o		
thick	NOUN	et		
thyme	NOUN	ol		
tierce	NOUN	el		
tine	NOUN	oid		
tonsilla	NOUN	itis		
trump	NOUN	et		
tub	NOUN	y		
tubercle	NOUN	ulate		
type	NOUN	o		
ultima	NOUN	ate		
vagus	NOUN	al		
vase	NOUN	iform		
venter	NOUN	ral		
wake	NOUN	en		
weld	VERB	ment		
whack	NOUN	o		
wrist	NOUN	let		
yaw	NOUN	s		
zone	NOUN	ula		

Appendix 55

Iterative suffixation analysis: input and output

Input: 2nd. secondary suffix set as ordered by the optimal heuristic

e	ight	ch	ar	ough	id	ow	ing
ook	ck	en	ss	t	el	ail	a
ouse	am	eed	our	oof	ino	ake	sh
eep	eek	ill	ack	ort	ailor	aw	ood
ast	low	iii	uff	ave	ink	ense	ock
ark	allow	ng	out	ther	arrow	il	ope
ump	owel	ash	eak	viii	aste	fish	aze
llow	orm	ank	ound	ign	asting	ext	xxv
oodoo	and	at	oot	or	ophyte	ob	h
ght	l	lock	eau	k	ram	old	d
ish	owl	arp	own	end	ac	illa	ore
aboo	rawl	unch	ass	it	ot	que	appa
ensor	weed	ame	ear	est	re	iff	wort
ouch	ebibit	ebibyte	iv	ap	tch	hirr	ierce
rowning	ern	xvi	xvii	xviii	atch	ick	ingo
arch	asp	unnel	each	ff	ome	op	tern
alm	raft	ad	eat	ead	ife	inge	ilt
orrhea	awk	arina	onym	ridge	alif	ealth	innow
occi	oncho	oplasm	rmaid	hyme	ndue	ulse	alve
amba	abbala	abbalah	ackbut	adderwort	adre	aggot	ahertz
airn	alanga	aliph	alpac	ampong	anana	ankeen	ansom
antra	apir	apote	arfare	arotid	arrot	arry	artridge
asbah	ascara	atchel	attail	aurel	avior	aviour	awp
earest	eckon	edick	edlar	edwood	eethe	ervid	escue
haddar	herefore	hittimwood	ickshaw	ilbert	illoma	ippo	irasol

Output: Results obtained with 2nd. secondary suffix set as ordered by the optimal heuristic

Original word	Original POS	Identified root	Root POS	Relation type
acantha	NOUN	acanthus	NOUN	MASCULINE
acneiform	ADJECTIVE	acne	NOUN	RESEMBLEDBY
aculea	NOUN	aculeus	NOUN	MASCULINE
agenda	NOUN	agendum	NOUN	SINGULAR
albuminoid	NOUN	albumin	NOUN	RESEMBLEDBY
alienor	NOUN	alien	VERB	ROLE
alumina	NOUN	aluminum	NOUN	SINGULAR
ampullar	ADJECTIVE	ampul	NOUN	PERTAINYM
amyloid	NOUN	amyl	NOUN	RESEMBLEDBY
amyloid	ADJECTIVE	amyl	NOUN	RESEMBLEDBY
anima	NOUN	animus	NOUN	MASCULINE
arboriform	ADJECTIVE	arbor	NOUN	RESEMBLEDBY
armilla	NOUN	arm	NOUN	FULLSIZE
armor	NOUN	arm	VERB	ROLE
astragalar	ADJECTIVE	astragal	NOUN	PERTAINYM
bailor	NOUN	bail	VERB	ROLE
barbel	NOUN	barb	NOUN	FULLSIZE
bethel	NOUN	beth	NOUN	FULLSIZE
bitumenoid	ADJECTIVE	bitumen	NOUN	RESEMBLEDBY

Original word	Original POS	Identified root	Root POS	Relation type
bulbar	ADJECTIVE	bulb	NOUN	PERTAINYM
bulbil	NOUN	bulb	NOUN	FULLSIZE
candelabra	NOUN	candelabrum	NOUN	SINGULAR
canella	NOUN	cane	NOUN	FULLSIZE
carbonyl	ADJECTIVE	carbon	NOUN	PERTAINYM
casquetel	NOUN	casquet	NOUN	FULLSIZE
chiasma	NOUN	chiasmus	NOUN	MASCULINE
christella	NOUN	christ	NOUN	FULLSIZE
cisterna	NOUN	cistern	NOUN	MASCULINE
clad	ADJECTIVE	clothe	VERB	VERB_SOURCE
clangor	NOUN	clang	VERB	ROLE
cockerel	NOUN	cocker	NOUN	FULLSIZE
colonel	NOUN	colon	NOUN	FULLSIZE
columnar	ADJECTIVE	column	NOUN	PERTAINYM
columniform	ADJECTIVE	column	NOUN	RESEMBLEDBY
cornea	NOUN	corneum	NOUN	SINGULAR
counsellor	NOUN	counsel	VERB	ROLE
counselor	NOUN	counsel	VERB	ROLE
ctenoid	ADJECTIVE	ctene	NOUN	RESEMBLEDBY
cubiform	ADJECTIVE	cube	NOUN	RESEMBLEDBY
cuboid	NOUN	cube	NOUN	RESEMBLEDBY
cuboid	ADJECTIVE	cube	NOUN	RESEMBLEDBY
cuneiform	NOUN	cuneiform	ADJECTIVE	ATTRIBUTE_VALUE
data	NOUN	datum	NOUN	SINGULAR
drunk	NOUN	drink	VERB	VERBSOURCE_OF_GERUND
drunk	ADJECTIVE	drink	VERB	VERB_SOURCE
dura	NOUN	durum	NOUN	SINGULAR
error	NOUN	err	VERB	ROLE
factoid	NOUN	fact	NOUN	RESEMBLEDBY
facula	NOUN	face	NOUN	FULLSIZE
fauna	NOUN	faun	NOUN	MASCULINE
flexor	NOUN	flex	VERB	ROLE
fluid	ADJECTIVE	flue	NOUN	QUALIFIED
folderol	NOUN	folder	NOUN	SUBSTANCE_HOLONYM
fulfill	VERB	fulfil	VERB	SYNONYM
fusiform	ADJECTIVE	fuse	NOUN	RESEMBLEDBY
fusil	NOUN	fuse	NOUN	FULLSIZE
gentianella	NOUN	gentian	NOUN	FULLSIZE
gingerol	NOUN	ginger	NOUN	SUBSTANCE_HOLONYM
gladiola	NOUN	gladiolus	NOUN	MASCULINE
governor	NOUN	govern	VERB	ROLE
gunnel	NOUN	gun	NOUN	FULLSIZE
held	ADJECTIVE	hold	VERB	VERB_SOURCE
hostel	NOUN	host	NOUN	FULLSIZE
humanoid	NOUN	human	NOUN	RESEMBLEDBY
jailor	NOUN	jail	VERB	ROLE
javelina	NOUN	javelin	NOUN	MASCULINE
laid	ADJECTIVE	lay	VERB	VERB_SOURCE
legionella	NOUN	legion	NOUN	FULLSIZE
liar	NOUN	lie	NOUN	HOME
linear	ADJECTIVE	line	NOUN	PERTAINYM

Original word	Original POS	Identified root	Root POS	Relation type
lobar	ADJECTIVE	lobe	NOUN	PERTAINYM
lurid	ADJECTIVE	lure	NOUN	QUALIFIED
ma	NOUN	mum	NOUN	SINGULAR
meteoroid	NOUN	meteor	NOUN	RESEMBLEDBY
mucinoid	ADJECTIVE	mucin	NOUN	RESEMBLEDBY
muscatel	NOUN	muscat	NOUN	FULLSIZE
neutrino	NOUN	neutron	NOUN	FULLSIZE
paid	ADJECTIVE	pay	VERB	VERB_SOURCE
palmar	ADJECTIVE	palm	NOUN	PERTAINYM
persona	NOUN	person	NOUN	MASCULINE
personnel	NOUN	person	NOUN	FULLSIZE
petaloid	ADJECTIVE	petal	NOUN	RESEMBLEDBY
pickerel	NOUN	picker	NOUN	FULLSIZE
planar	ADJECTIVE	plane	NOUN	PERTAINYM
planetoid	NOUN	planet	NOUN	RESEMBLEDBY
planula	NOUN	plane	NOUN	FULLSIZE
primula	NOUN	prime	NOUN	FULLSIZE
prismoid	NOUN	prism	NOUN	RESEMBLEDBY
razor	NOUN	raze	VERB	ROLE
resinoid	NOUN	resin	NOUN	RESEMBLEDBY
rhea	NOUN	rheum	NOUN	SINGULAR
rhomboid	NOUN	rhomb	NOUN	RESEMBLEDBY
rhomboid	ADJECTIVE	rhomb	NOUN	RESEMBLEDBY
rosilla	NOUN	rose	NOUN	FULLSIZE
roundel	NOUN	round	NOUN	FULLSIZE
said	ADJECTIVE	say	VERB	VERB_SOURCE
sailor	NOUN	sail	VERB	ROLE
salmonella	NOUN	salmon	NOUN	FULLSIZE
salmonid	NOUN	salmon	ADJECTIVE	QUALIFYING
salverform	ADJECTIVE	salver	NOUN	RESEMBLEDBY
salvor	NOUN	salve	VERB	ROLE
scalar	NOUN	scale	NOUN	HOME
scalar	ADJECTIVE	scale	NOUN	PERTAINYM
sensor	NOUN	sense	VERB	ROLE
settlor	NOUN	settle	VERB	ROLE
shod	ADJECTIVE	shoe	VERB	VERB_SOURCE
sinusoid	NOUN	sinus	NOUN	RESEMBLEDBY
sold	ADJECTIVE	sell	VERB	VERB_SOURCE
spheroid	NOUN	sphere	NOUN	RESEMBLEDBY
succuba	NOUN	succubus	NOUN	MASCULINE
sunk	ADJECTIVE	sink	VERB	VERB_SOURCE
tabloid	NOUN	table	NOUN	RESEMBLEDBY
tensor	NOUN	tense	VERB	ROLE
tercel	NOUN	terce	NOUN	FULLSIZE
thymol	NOUN	thyme	NOUN	SUBSTANCE_HOLONYM
tiercel	NOUN	tierce	NOUN	FULLSIZE
tineoid	NOUN	tine	NOUN	RESEMBLEDBY
toroid	NOUN	tore	NOUN	RESEMBLEDBY
umbellar	ADJECTIVE	umbel	NOUN	PERTAINYM
umbelliform	ADJECTIVE	umbel	NOUN	RESEMBLEDBY
vaccina	NOUN	vaccinum	NOUN	SINGULAR

Original word	Original POS	Identified root	Root POS	Relation type
vasiform	ADJECTIVE	vase	NOUN	RESEMBLEDBY
vendor	NOUN	vend	VERB	ROLE
virusoid	NOUN	virus	NOUN	RESEMBLEDBY
zonula	NOUN	zone	NOUN	FULLSIZE

Input: 3rd. secondary suffix set as ordered by the default heuristic

e	ng	id	a	ck	t	ing	ar
el	ch	ss	d	ght	ow	en	l
wort	h	ort	ight	sh	lla	la	ish
re	se	or	am	oid	k	r	orm
o	il	ll	ff	iform	form	eed	th
che	saur	ur	osaur	ack	st	raph	scope
ook	oscope	illa	ent	graph	nd	ac	rn
ograph	ock	ood	ouse	rt	ore	aph	ail
at	tch	our	ogram	ast	ough	ope	cope
wood	op	gram	oma	fish	ot	rm	ass
m	om	ake	g	and	ill	ad	ocyte
phyte	yte	it	ma	asm	ead	est	te
ino	ra	own	ugh	llo	ram	out	nch
ophyte	llow	bird	ase	use	ick	que	n
ol	na	ern	ave	aw	eak	ark	eau
nk	dge	here	p	ina	ign	oot	low
mp	ound	ula	rrow	ogen	erwort	sphere	eep
orrhea	ile	ge	gue	ica	le	ella	ank
ophore	nge	smith	iii	weed	head	oof	tz
ome	arp	ith	ah	i	ird	ord	illo
ash	lock	ump	phore	to	type	ew	me
ink	otype	od	esce	ap	dom	the	root
uff	row	ime	end	osphere	pe	aur	eek
aste	ield	old	ther	iece	inase	awk	bibyte
troke	inogen	osome	iff	phere	ense	chi	aft

Output: Results obtained with 3rd. secondary suffix set as ordered by the default heuristic

Original word	Original POS	Identified root	Root POS	Relation type
ani	NOUN	anus	NOUN	SINGULAR
beano	NOUN	bean	NOUN	ROOT
billyo	NOUN	billy	NOUN	ROOT
boredom	NOUN	bore	NOUN	POSSESSOR_OF_ATTRIBUTE
bravo	NOUN	brave	NOUN	ROOT
canto	NOUN	cant	NOUN	ROOT
cocci	NOUN	coccus	NOUN	SINGULAR
condom	NOUN	con	NOUN	POSSESSOR_OF_ATTRIBUTE
dug	NOUN	dig	VERB	VERBSOURCE_OF_GERUND
dukedom	NOUN	duke	NOUN	POSSESSOR_OF_ATTRIBUTE
earldom	NOUN	earl	NOUN	POSSESSOR_OF_ATTRIBUTE
fandom	NOUN	fan	NOUN	POSSESSOR_OF_ATTRIBUTE
fiefdom	NOUN	fief	NOUN	POSSESSOR_OF_ATTRIBUTE
filmdom	NOUN	film	NOUN	POSSESSOR_OF_ATTRIBUTE
flamingo	NOUN	flaming	NOUN	ROOT
freedom	NOUN	free	NOUN	POSSESSOR_OF_ATTRIBUTE
gangdom	NOUN	gang	NOUN	POSSESSOR_OF_ATTRIBUTE

Original word	Original POS	Identified root	Root POS	Relation type
gyro	NOUN	gyre	NOUN	ROOT
kingdom	NOUN	king	NOUN	POSSESSOR_OF_ATTRIBUTE
loti	NOUN	lotus	NOUN	SINGULAR
magneto	NOUN	magnet	NOUN	ROOT
martyrdom	NOUN	martyr	NOUN	POSSESSOR_OF_ATTRIBUTE
medico	NOUN	medic	NOUN	ROOT
memo	NOUN	meme	NOUN	ROOT
mimeo	NOUN	mime	NOUN	ROOT
mini	NOUN	minus	NOUN	SINGULAR
nardoo	NOUN	nardo	NOUN	ROOT
nympho	NOUN	nymph	NOUN	ROOT
pi	NOUN	pus	NOUN	SINGULAR
pinko	NOUN	pink	NOUN	ROOT
porno	NOUN	porn	NOUN	ROOT
primo	NOUN	prime	NOUN	ROOT
princedom	NOUN	prince	NOUN	POSSESSOR_OF_ATTRIBUTE
pruno	NOUN	prune	NOUN	ROOT
pseudo	NOUN	pseud	NOUN	ROOT
secondo	NOUN	second	NOUN	ROOT
serfdom	NOUN	serf	NOUN	POSSESSOR_OF_ATTRIBUTE
sheikdom	NOUN	sheik	NOUN	POSSESSOR_OF_ATTRIBUTE
sheikhdom	NOUN	sheikh	NOUN	POSSESSOR_OF_ATTRIBUTE
slew	NOUN	slay	VERB	VERBSOURCE_OF_GERUND
sodom	NOUN	so	NOUN	POSSESSOR_OF_ATTRIBUTE
staphylococci	NOUN	staphylococcus	NOUN	SINGULAR
stardom	NOUN	star	NOUN	POSSESSOR_OF_ATTRIBUTE
tamarindo	NOUN	tamarind	NOUN	ROOT
tenno	NOUN	ten	NOUN	ROOT
thralldom	NOUN	thrall	NOUN	POSSESSOR_OF_ATTRIBUTE
two	ADJECTIVE	second	ADJECTIVE	ADJECTIVE_SOURCE
typo	NOUN	type	NOUN	ROOT
whacko	NOUN	whack	NOUN	ROOT
whoredom	NOUN	whore	NOUN	POSSESSOR_OF_ATTRIBUTE
yobbo	NOUN	yob	NOUN	ROOT
yobo	NOUN	yob	NOUN	ROOT

Input: 4th. secondary suffix set as ordered by the optimal heuristic

e	ight	ii	ch	ough	ow	ook	ck
t	ing	ss	en	am	ouse	eed	ake
sh	eep	eek	ack	ort	ood	ast	iii
ink	our	uff	ave	ense	oof	ock	ark
aw	allow	ng	ther	arrow	low	ope	h
k	ump	ash	eak	viii	aste	fish	out
ank	llow	nd	ound	ign	asting	ext	xxv
and	at	oot	ophyte	aze	ob	ght	lock
eau	ram	owl	arp	own	ore	rawl	unch
ass	ur	ot	que	weed	old	oom	est
end	iff	ouch	ebibit	ebibyte	iv	ap	hirr
ierce	rowning	ern	xvi	xvii	xviii	atch	ick
ish	it	arch	asp	each	ff	ome	ame
od	op	tern	alm	raft	eat	ife	ield
inge	ilt	ac	awk	onym	ridge	alif	ealth

innow	oplasm	hyme	ulse	alve	abbalah	ackbut	adderwort
adre	aggot	ahertz	airn	aliph	alpac	ampong	ankeen
ansom	apir	apote	arfare	arrot	arry	artridge	asbah
aviour	awp	earest	eckon	edick	edwood	eethe	escue
herefore	hittimwood	ickshaw	ilbert	ivot	lamour	niseed	ogwood
olliwog	olograph	oluble	ootle	otshot	ouffe	umquat	urbot
urrajong	urrawong	ill	tch	oscope	wood	re	usk
ll	ird	awl	oke	omb	row	ograph	ew
amp	ase	oupe	arnish	ittern	xxi	xxii	xxiii
xxiv	xxvi	xxvii	xxviii	che	iece	ogue	se

Output: No results were obtained with 4th. secondary suffix set as ordered by the optimal heuristic

Input: 5th. secondary suffix set as ordered by the default heuristic

e	t	ng	ck	ing	ch	ss	h
ur	ght	ow	en	wort	ort	ight	k
sh	nd	am	ish	re	se	ll	ff
d	eed	th	che	saur	osaur	ack	st
raph	ii	scope	ook	oscope	ent	graph	ac
rn	g	ograph	ock	ood	ouse	rt	ore
aph	at	tch	our	ogram	ast	n	ough
ope	cope	wood	op	gram	fish	ot	m
p	ass	ake	and	ocyte	phyte	od	yte
it	asm	est	te	own	ugh	ram	out
nch	ophyte	llow	bird	ase	use	ick	que
ern	ave	eak	ark	eau	aw	dge	here
nk	ign	oot	low	mp	ound	rrow	ogen
erwort	ir	sphere	eep	ile	ge	gue	ank
le	ophore	iii	ill	nge	om	smith	weed
oof	tz	ome	arp	ith	ah	ird	ord
ash	lock	ump	oom	phore	ink	type	me
otype	rd	r	esce	ap	ew	ed	ld
the	root	uff	ield	row	ime	end	osphere
pe	aur	eek	aste	ther	iece	inase	awk
bibyte	troke	inogen	osome	iff	phere	ense	aft
old	arch	ain	awl	ire	und	orn	spore
ob	l	er	ut	ife	wright	ere	ogue
bibit	ear	ospore	trix	ong	ue	cyte	tern
house	arrow	otte	hore	carp	allow	owl	alk

Output: No results were obtained with 5th. secondary suffix set as ordered by the default heuristic

Appendix 56

Iterative prefixation analysis: input and output

Input: 2nd. secondary prefix set as ordered by the optimal heuristic

s	c	qu	lxx	squ	b	t	st	ha	p
ro	fl	lxxx	ca	fla	sc	f	lo	co	gr
th	asco	bathyscap	handi	bo	sh	gro	ho	sno	pro
ch	g	xx	ta	ra	xxx	ba	sp	la	ya
sheat	ma	da	cra	br	whi	glo	l	cr	po
slo	me	har	qui	myria	seismo	absint	cantalou	chemis	chilias
chrono	clxx	cusha	e'e	fantas	highfaluti	idio	leitmoti	mave	megil
mollus	mulc	petti	planocon	pleonas	pontif	ravigot	regim	roentgeno	sapien
satisf	serap	smidg	somato	somewh	teet	thingama	thingma	thinguma	thrus
tomba	turbo	yashma	thro	sla	ri	thr	dra	for	di
holo	m	ski	sca	ove	bur	ne	d	squa	cro
tama	blo	twi	swi	kno	tr	snoo	swa	va	arti
cove	ideo	meshugg	sporophy	susp	bene	jo	zi	fi	fo
gra	bar	ga	pl	meri	abys	alky	apac	dupl	fello
polly	salaa	shallo	skul	velou	wallo	wreat	flo	wi	bla
sha	shel	squir	scra	shi	h	che	no	hal	ja
de	cal	gna	blan	w	le	cla	wa	na	dr
wor	schno	telo	tur	tra	tro	sil	dis	bu	sto
war	crum	ple	bri	por	ver	brea	guil	spiro	clo
cur	sho	bl	ka	ve	car	chur	spor	pr	he
tu	mus	yo	cha	wel	cor	to	pu	mo	spri
sch	qua	bathys	meshug	olig	schti	sporoph	budg	canta	coho
hygro	kara	kha	roentge	secreta	shall	where	grea	aard	alba
angeli	ankylos	archit	aspar	aya	baili	belda	bolloc	boton	burea
calpa	carpo	challeng	chauffeu	chutzpa	clado	claus	coiff	conidio	corte
cring	danseu	devoi	equi	gametoph	goitr	golliwo	habi	hier	hologra
ibid	ideogra	kaffi	khali	kibbut	kolkho	kurra	lentis	lxvi	lyso
mackin	marqu	nabo	nomogra	nudni	oosp	ostraco	pedago	phala	pheno
phonogra	pillo	pinnati	piro	pizza	pterido	putref	sandara	schmal	seismogra
shella	shno	sidero	silve	skiagra	sleig	soign	sonogra	spher	spirogra
spong	styra	sulfu	suspen	syrin	tachogra	tchotchk	telomer	twili	vapo
virt	wron	xanthophy	xcvi	xlvi	xxvi	thor	xxxi	xxxv	clim
prim	snar	allo	centro	glea	massi	miao	mont	phlo	sara
sco	fr	a	lx	scr	re	shir	lin	suc	thin
wh	hoo	cho	spo	ran	du	slu	leas	plum	syn
or	al	sta	uro	what	fe	ser	se	aga	mor
cas	arche	pico	pila	bra	her	rou	sa	cus	ste
squi	za	sna	scal	whel	glu	fra	fro	she	shti
stor	brus	screa	smar	swea	swee	thum	ni	gl	tri
cre	ar	spi	wal	pre	thi	benef	fond	breat	ear
heli	kur	lxxxi	lxxxv	broo	cree	roo	duc	spir	mal
gri	stra	whe	wo	bea	blin	cit	ther	nic	gol
el	tuss	wri	r	trou	stri	flu	flam	ru	crus
ju	medi	star	acol	ambi	amon	auro	barbe	benefi	branc
breath	cair	carib	centim	dall	gyno	handic	hicc	homb	indi
kope	ligh	lxxvi	lxxxvi	muta	neig	neve	oce	orang	philo
proteo	strang	xxxvi	xc	spur	whor	fres	orac	pinc	strea
vi	bal	bas	cer	lou	pla	cu	pil	ze	ur
shor	lea	pur	do	ora	grap	yaw	sporo	bul	swo
ven	seri	tera	vers	rus	smi	pra	lu	mar	k

Output: Results obtained with 2nd. secondary prefix set as ordered by the optimal heuristic

Original word	Prefix	Stem	Original word	Prefix	Stem
ambient	ambi	ient	hygroscope	hygro	scope
archeopteryx	arche	pteryx	ideogram	ideo	gram
archespore	arche	spore	ideograph	ideo	graph
archetype	arche	type	ideologue	ideo	logue
artifact	arti	fact	karaoke	kara	oke
artiste	arti	ste	lysosome	lyso	some

Original word	Prefix	Stem	Original word	Prefix	Stem
benedick	bene	dick	lysozyme	lyso	zyme
benefact	bene	fact	maladroit	mal	adroit
beneficent	bene	ficent	malaise	mal	aise
benefit	bene	fit	malaprop	mal	aprop
carpophore	carpo	phore	maleficent	mal	eficent
carpospore	carpo	spore	malign	mal	ign
chronograph	chrono	graph	malnourish	mal	nourish
chronoscope	chrono	scope	malodour	mal	odour
duplex	dupl	ex	maltreat	mal	treat
flambe	flam	be	mericarp	meri	carp
flambeau	flam	beau	meristem	meri	stem
fondue	fond	ue	montane	mont	ane
halophyte	hal	phyte	mutafacient	muta	facient
heliac	heli	ac	mutagen	muta	gen
holocaust	holo	caust	myriagram	myria	gram
hologram	holo	gram	myriametre	myria	metre
holograph	holo	graph	myriapod	myria	pod
holonym	holo	onym	oligarch	olig	arch
holophyte	holo	phyte	oligochaete	olig	chaete
holotype	holo	type	oligoclase	olig	clase
hygrodeik	hygro	deik	oligodendrocyte	olig	dendrocyte
hygrophyte	hygro	phyte	phenoplast	pheno	plast
			phenotype	pheno	type
picometre	pico	metre			
picosecond	pico	second			
picovolt	pico	volt			
pteridophyte	pterido	phyte			
pteridosperm	pterido	sperm			
retrieve	re	trieve			
scalene	scal	ene			
somatosense	somato	sense			
somatotype	somato	type			
spherocyte	spher	cyte			
spirit	spir	it			
spirochaete	spir	chaete			
spirochete	spir	chete			
spirogram	spir	gram			
spirograph	spir	graph			
spongioblast	spong	ioblast			
sporangiophore	spor	angiophore			
sporocarp	spor	carp			
sporophore	spor	phore			
sporophyl	spor	phyl			
sporophyll	spor	phyll			
sporophyte	spor	phyte			
syringe	syrin	ge			
telomerase	telo	merase			
telomere	telo	mere			
telophase	telo	phase			
theropod	ther	pod			
urease	ur	ease			

Original word	Prefix	Stem	Original word	Prefix	Stem
urobilinogen	ur	bilinogen			
urochord	ur	chord			
urokinase	ur	kinase			
urolith	ur	lith			

Input: 3rd. secondary prefix set as ordered by the optimal heuristic

s	c	qu	lxx	squ	b	t	st	ha	p
ro	fl	lxxx	ca	sc	lo	f	co	gr	fla
th	asco	bathyscap	handi	bo	sh	gro	sno	pro	g
ho	xx	ch	ta	ra	xxx	ba	la	ya	sheat
da	ma	cra	br	whi	glo	sp	l	po	me
cr	har	slo	qui	seismo	absint	cantalou	chemis	chilias	clxx
cusha	e'e	fantas	highfaluti	idio	leitmoti	mave	megil	mollus	mulc
petti	planocon	pleonas	pontif	ravigot	regim	roentgeno	sapien	satisf	serap
smidg	somewh	teet	thingama	thingma	thinguma	thrus	tomba	turbo	yashma
thro	sla	thr	for	ri	dra	di	ski	m	d
ove	bur	ne	squa	cro	tama	sca	blo	twi	swi
tr	kno	snoo	swa	va	cove	meshugg	susp	jo	fi
zi	gra	bar	fo	pl	flo	ga	abys	alky	apac
fello	polly	salaa	shallo	skul	velou	wallo	wreat	wi	bla
sha	shel	che	squir	scra	shi	cal	no	w	de
ja	gna	blan	h	le	cla	dr	wa	na	wor
schno	tur	hal	tra	tro	sil	bu	dis	sto	war
crum	ple	bri	por	ver	brea	guil	clo	cur	mus
bl	sho	pr	he	ve	chur	tu	cha	mo	ka
to	yo	hoo	wel	cor	pu	car	sch	spri	qua
bathys	meshug	schti	budg	canta	coho	kha	roentge	secreta	shall
where	grea	aard	alba	angeli	ankylos	archit	aspar	aya	baili
belda	bolloc	boton	burea	calpa	challeng	chauffeu	chrom	chutzpa	clado
claus	coiff	conidio	corte	cring	danseu	devoi	equi	gametoph	goitr
golliwo	habi	hier	ibid	kaffi	kara	khali	kibbut	kolkho	kurra
lentis	lxvi	mackin	marqu	nabo	nomogra	nudni	oosp	ostraco	pedago
phala	phonogra	pillo	pinnati	piro	pizza	ptero	putref	sandara	schmal
seismogra	shella	shno	sidero	silve	skiagra	sleig	soign	sonogra	styra
sulfu	suspen	tachogra	tchotchk	twili	vapo	virt	wron	xanthophy	xcvi
xlvi	xxvi	thor	xxxi	xxxv	clim	prim	snar	syn	allo
centro	glea	massi	miao	phlo	sara	sco	fr	lx	scr
a	shir	re	lin	suc	cho	thin	wh	or	ran
al	slu	leas	plum	fe	sta	what	du	se	mor
cas	her	ser	sa	aga	pila	bra	rou	cus	ste
squi	za	sna	whel	glu	fra	fro	she	shti	bea
stor	brus	screa	smar	swea	swee	thum	ni	gl	tri
cre	duc	wal	thi	pre	breat	ear	kur	lxxxi	lxxxv
broo	cree	roo	gri	stra	whe	wo	blin	cit	nic
gol	el	flu	r	tuss	scal	wri	trou	pil	stri
ru	crus	ar	ju	medi	star	acol	amon	auro	barbe
branc	breath	cair	carib	centim	dall	fond	gyno	handic	hicc
homb	indi	kope	ligh	lxxvi	lxxxvi	neig	neve	oce	orang
philo	proteo	strang	xxxvi	xc	spur	whor	fres	orac	pinc
strea	vi	bal	bas	spi	cer	cu	lou	pla	mar
ze	shor	lea	pur	do	ora	grap	yaw	bul	swo
ven	seri	tera	vers	rus	lu	sou	smi	pra	k
wha	carac	giga	mish	over	ribo	tropo	ber	scri	bel
cour	slee	ther	num	ble	plas	ama	gi	cle	chee
sal	scar	heli	horo	hors	pran	shriv	smit	squar	veno
spo	char	ker	min	dir	dru	wil	ter	tus	hu

Output: Results obtained with 3rd. secondary suffix set as ordered by the default heuristic

Original word	Prefix	Stem
gigabit	giga	bit
gigabyte	giga	byte
gigahertz	giga	hertz
horologe	horo	loge
horoscope	horo	scope

minuend	min	uend
plasmacyte	plas	macyte
plasminogen	plas	minogen
plastique	plas	tique
pterodactyl	ptero	dactyl
pterosaur	ptero	saur

Input: 4th. secondary prefix set as ordered by the optimal heuristic

s	c	qu	lxx	squ	b	t	st	ha	p
ro	fl	lxxx	ca	sc	f	lo	co	gr	fla
th	asco	bathyscap	handi	bo	sh	gro	sno	pro	g
xx	ch	ta	ho	ra	xxx	ba	la	ya	sheat
da	ma	cra	br	whi	glo	po	sp	l	me
cr	har	slo	qui	seismo	absint	cantalou	chemis	chilias	clxx
cusha	e'e	fantas	highfaluti	idio	leitmoti	mave	megil	mollus	mulc
petti	planocon	pleonas	pontif	ravigot	regim	roentgeno	sapien	satisf	serap
smidg	somewh	teet	thingama	thingma	thinguma	thrus	tomba	turbo	yashma
thro	sla	thr	for	ri	dra	di	ski	m	d
ove	bur	ne	squa	cro	tama	sca	blo	twi	swi
tr	kno	snoo	swa	va	cove	meshugg	susp	jo	fi
zi	gra	bar	fo	ga	flo	abys	alky	apac	fello
polly	salaa	shallo	skul	velou	wallo	wreat	wi	bla	sha
shel	pl	che	squir	scra	shi	cal	w	no	de
ja	gna	blan	le	h	cla	ple	dr	wa	na
wor	schno	tur	hal	tra	tro	sil	bu	dis	sto
war	crum	bri	por	ver	brea	guil	clo	cur	mus
pr	bl	sho	he	hoo	ve	chur	tu	cha	mo
pu	ka	to	yo	wel	cor	car	sch	spri	qua
bathys	meshug	schti	budg	canta	coho	kha	roentge	secreta	shall
where	grea	aard	alba	angeli	ankylos	archit	aspar	aya	baili
belda	bolloc	boton	burea	calpa	challeng	chauffeu	chrom	chutzpa	clado
claus	coiff	conidio	corte	cring	danseu	devoi	equi	gametoph	goitr
golliwo	habi	hier	ibid	kaffi	kara	khali	kibbut	kolkho	kurra
lentis	lxvi	mackin	marqu	nabo	nomogra	nudni	oosp	ostraco	pedago
phala	phonogra	pillo	pinnati	piro	pizza	putref	sandara	schmal	seismogra
shella	shno	sidero	silve	skiagra	sleig	soign	sonogra	styra	sulfu
suspen	tachogra	tchotchk	twili	vapo	virt	wron	xanthophy	xcvi	xlvi
xxvi	thor	xxxi	xxxv	clim	prim	snar	syn	allo	centro
glea	massi	miao	phlo	sara	sco	fr	lx	a	scr
shir	re	lin	suc	cho	thin	wh	or	ran	al
slu	leas	plum	fe	sta	what	du	se	mor	cas
her	ser	sa	aga	pila	bra	rou	cus	ste	squi
za	sna	whel	glu	fra	fro	gl	she	shti	bea
stor	brus	hors	screa	smar	swea	swee	thum	ni	tri
cre	duc	wal	thi	pre	breat	ear	kur	lxxxi	lxxxv
broo	cree	roo	gri	stra	whe	wo	blin	cit	nic
gol	el	r	flu	tuss	scal	wri	trou	pil	stri
ru	crus	ar	ju	medi	star	acol	amon	auro	barbe
branc	breath	cair	carib	centim	dall	fond	gyno	handic	hicc
homb	indi	kope	ligh	lxxvi	lxxxvi	neig	neve	oce	orang
philo	proteo	strang	xxxvi	xc	spur	whor	fres	orac	pinc
strea	vi	bal	bas	spi	cer	cu	lou	mar	ze
shor	lea	pur	do	ora	grap	yaw	bul	swo	ven
seri	tera	vers	rus	lu	sou	smi	pra	k	wha
carac	mish	over	ribo	tropo	ber	scri	bel	cour	slee
ther	num	ble	ama	cle	chee	pla	sal	plo	scar
heli	pran	shriv	smit	squar	veno	spo	char	ker	dir
dru	wil	hu	ter	tus	blit	sni	gros	pe	lim

Output: No results were obtained with 4th. secondary prefix set as ordered by the optimal heuristic

Input: 5th. secondary prefix set as ordered by the default heuristic

car	cent	for	ver	bar	in	thing	bur	ove	an
asco	coel	melan	bathys	meshug	thin	gen	har	cal	ter
tuss	or	al	ar	cur	tama	cen	obe	budg	coho
ostr	canta	handi	mujah	prote	shall	techn	where	gameto	seismo
roentge	secreta	bathyscap	ser	est	arch	medi	tamar	mor	mar

ran	ball	bors	cor	dis	guil	some	oxi	ult	cove
fell	hist	lact	phil	ravi	susp	chall	sheat	meshugg	am
her	tur	bath	war	ama	el	aqu	aya	e'e	aard
alba	aris	azed	bo's	cycl	equi	gymn	habi	hier	ibid
idio	kara	loll	mave	mulc	nabo	nebb	neph	oosp	piro
roll	teet	vapo	vigo	virt	wron	aspar	baili	baksh	belda
boton	burea	calpa	carca	chitt	chrom	clado	claus	coiff	corte
costu	cring	curra	cusha	devoi	febri	fissi	gibib	goitr	kaffi
khali	kurra	leuco	lique	magni	marqu	mebib	megil	nudni	pachy
pebib	petab	petti	phala	pillo	pizza	regim	sauer	serap	shill
shitt	silve	sleig	smidg	soign	styra	sulfu	tebib	thrus	tomba
turbo	twili	yobib	zebib	absint	angeli	archit	bolloc	budger	carrag
chemis	chlamy	danseu	fantas	fibrin	kibbut	kolkho	lentis	mackin	mollus
pedago	phosph	pontif	putref	sapien	satisf	schmal	shella	sidero	sinist
somewh	sprech	sterco	suspen	tovari	yashma	yottab	zettab	ankylos	chilias
chutzpa	conidio	golliwo	nomogra	ostraco	pinnati	pleonas	ravigot	sandara	skiagra
sonogra	thingma	cantalou	challeng	chauffeu	gametoph	leitmoti	phonogra	planocon	spermato
tachogra	tchotchk	thingama	thinguma	ribonucle	roentgeno	seismogra	xanthophy	ballistoca	highfaluti
centr	crum	hall	lan	hal	ora	tam	wel	long	mish
over	ribo	carac	tropo	wor	chur	what	mus	sil	gol
por	ber	bat	shel	blan	men	cer	ava	cach	kibb
kibi	oran	pinn	poll	sati	thor	wall	val	mas	cir
cit	blin	lang	kin	vel	ven	sal	bul	aug	int
bil	oce	usu	abys	acol	alky	amon	anne	apac	auro
buck	cair	dall	elas	fond	gyno	hect	hicc	homb	hyal
indi	keto	kope	ligh	litt	neig	neve	ninj	oxid	siam
skul	sync	tume	volu	yogh	barbe	branc	carib	champ	fello
kibib	morph	orang	phant	philo	polly	quand	salaa	stoma	trave
velou	wallo	wreat	breath	centim	handic	proteo	shallo	strang	techno
mass	star	dan	lin	suc	chor	cas	tus	bill	kind
lent	moll	pila	sand	velo	squir	bor	trop	tac	seri
tera	vers	pil	res	arc	arg	fin	baro	scal	shir
min	aga	ear	kur	coll	larg	mani	phan	phon	resi
breat	centi	cel	char	pur	bal	bas	fur	ast	hel
kib	kit	len	ten	bon	lar	axi	ent	euc	eve
ima	oes	agai	allo	anim	anth	circ	hack	have	hemi
holl	madr	meag	napr	negl	nigh	noug	pali	remi	sara
suma	supe	supr	tast	weig	yarm	blint	carre	chang	coelo
creas	grand	graph	guill	langu	massi	shtic	terab	whirl	centro
melano	schtic	tamara	tamari	gyn	opa	syn	bulg	clim	geno
maca	prim	snar	spur	tach	whor	whir	kal	bir	bis
mel	mes	tar	fet	duc	per	tom	tor	pas	wal
som	cour	dist	leas	plum	sala	ther	bel	pin	gul
nar	cara	mol	as	mit	yar	gran	grap	cul	cus
dir	er	mac	mat	aby	zeb	blit	rang	whel	stran

Output: Results obtained with 5th. secondary prefix set as ordered by the default heuristic

Original word	Prefix	Stem	Original word	Prefix	Stem
animadvert	anim	advert	hectare	hect	are
aqueduct	aqu	educt	hemiepiphyte	hemi	epiphyte
aquilege	aqu	lege	hemisphere	hemi	sphere
architect	arch	tect	histaminase	hist	aminase
architrave	arch	trave	histiocyte	hist	iocyte
archosaur	arch	saur	histogram	hist	gram
augend	aug	end	ketoprofen	keto	profen
augur	aug	ur	ketorolac	keto	rolac
august	aug	ust	lactase	lact	ase
axile	axi	le	lactifuge	lact	fuge
ballast	ball	ast	lactogen	lact	gen
ballistocardiogram	ball	istocardiogram	leucocyte	leuco	cyte
ballistocardiograph	ball	istocardiograph	leucothoe	leuco	thoe
ballock	ball	ock	magnificent	magni	ficent
ballot	ball	ot	magniloquent	magni	loquent

Original word	Prefix	Stem	Original word	Prefix	Stem
batholith	bath	lith	melancholiac	melan	choliac
bathyscape	bath	scape	melanoblast	melan	blast
bathyscaph	bath	scaph	melanocyte	melan	cyte
bathyscaphe	bath	scaphe	mollusc	moll	usc
bathysphere	bath	sphere	mollusk	moll	usk
centrex	centr	ex	pachycephalosaur	pachy	cephalosaur
centrifuge	centr	fuge	pachyderm	pachy	derm
centromere	centr	mere	philologue	phil	logue
centrosome	centr	some	philomath	phil	math
choreograph	chor	ograph	phoneme	phon	eme
coelacanth	coel	acanth	phonogram	phon	gram
coeliac	coel	ac	phonograph	phon	graph
coelom	coel	om	phosphatase	phosph	atase
coelostat	coel	stat	phosphoresce	phosph	oresce
cyclamen	cycl	men	pinniped	pinn	ped
cycles/second	cycl	s/second	proteinase	prote	inase
febrifuge	febri	fuge	proteome	prote	ome
febrile	febri	ile	proteosome	prote	some
gendarme	gen	darme	stercobilinogen	sterco	bilinogen
genome	gen	ome	stercolith	sterco	lith
genotype	gen	type	supreme	supr	eme
gentle	gen	le	tachistoscope	tach	istoscope
grapheme	graph	eme	tachogram	tach	gram
graphospasm	graph	spasm	tachograph	tach	graph
gymnast	gymn	ast	technique	techn	ique
gymnosperm	gymn	sperm	technocrat	techn	crat
gynandromorph	gyn	andromorph	technophobe	techn	phobe
gynobase	gyn	base	trophoblast	trop	hoblast
gynophore	gyn	phore	troponym	trop	nym
tropopause	trop	pause			
troposphere	trop	sphere			

Input: 6th. secondary prefix set as ordered by the default heuristic

car	for	bar	ver	in	thing	bur	ove	an	asco
meshug	ter	thin	har	cal	tuss	al	or	cur	tama
obe	budg	coho	ostr	canta	handi	mujah	shall	where	gameto
seismo	roentge	secreta	ar	ser	mor	est	cent	medi	tamar
mar	ran	bors	cor	dis	her	guil	some	am	oxi
ult	cove	fell	ravi	susp	centi	chall	sheat	meshugg	tur
war	ama	el	lan	aya	e'e	aard	alba	aris	azed
bo's	equi	habi	hier	ibid	idio	kara	loll	mave	mulc
nabo	nebb	neph	oosp	piro	roll	teet	vapo	vigo	virt
wron	aspar	baili	baksh	belda	boton	burea	calpa	carca	chitt
chrom	clado	claus	coiff	corte	costu	cring	curra	cusha	devoi
fissi	gibib	goitr	kaffi	khali	kurra	lique	marqu	mebib	megil
nudni	pebib	petab	petti	phala	pillo	pizza	regim	sauer	serap
shill	shitt	silve	sleig	smidg	soign	styra	sulfu	tebib	thrus
tomba	turbo	twili	yobib	zebib	absint	angeli	bolloc	budger	carrag
chemis	chlamy	danseu	fantas	fibrin	kibbut	kolkho	lentis	mackin	pedago
pontif	putref	sapien	satisf	schmal	shella	sidero	sinist	somewh	sprech
suspen	tovari	yashma	yottab	zettab	ankylos	chilias	chutzpa	conidio	golliwo
nomogra	ostraco	pinnati	pleonas	ravigot	sandara	skiagra	sonogra	thingma	cantalou
challeng	chauffeu	gametoph	leitmoti	planocon	spermato	tchotchk	thingama	thinguma	ribonucle
roentgeno	seismogra	xanthophy	highfaluti	tam	crum	hall	hal	ora	wel
long	mish	over	ribo	carac	wor	chur	what	mus	sil
cer	gol	por	men	ber	shel	blan	ava	cach	kibb
kibi	oran	poll	sati	thor	wall	val	mas	cir	cit
blin	lang	kin	vel	ven	sal	bul	gen	int	bil

suc	oce	usu	abys	acol	alky	amon	anne	apac	auro
buck	cair	dall	elas	fond	hicc	homb	hyal	indi	kope
ligh	litt	neig	neve	ninj	oxid	siam	skul	supe	sync
tume	volu	yogh	barbe	branc	carib	champ	fello	kibib	morph
orang	phant	polly	quand	salaa	stoma	trave	velou	wallo	wreat
breath	centim	handic	shallo	strang	mass	arg	star	dan	lin
cas	tus	bill	kind	lent	pila	sand	velo	squir	cel
cen	bor	pil	bas	seri	tera	vers	res	fin	baro
scal	shir	min	aga	hel	ear	kur	coll	larg	mani
phan	resi	breat	char	pur	ten	len	fur	ast	lar
kib	kit	bon	mes	tar	fet	ent	euc	eve	ima
oes	agai	allo	anth	circ	hack	have	holl	madr	meag
napr	negl	nigh	noug	pali	remi	sara	suma	tast	weig
yarm	blint	carre	chang	chord	creas	grand	guill	langu	massi
shtic	terab	whirl	schtic	tamara	tamari	opa	syn	bulg	clim
maca	prim	snar	spur	whor	whir	kal	bir	bis	pas
duc	per	tom	tor	gran	wal	som	cour	dist	leas
plum	sala	ther	bel	gul	nar	cara	as	chor	mit
mac	mat	yar	cul	cus	dir	er	aby	zeb	blit
rang	whel	stran	pun	put	pos	air	ecr	pyr	tyr
brus	bunc	comf	dear	galo	geni	glit	gour	hors	intu
kali	knac	legi	peni	pinc	recu	riba	sabo	sacr	sens
smar	thum	weal	wild	borsc	borsh	hallu	scall	sprin	strob
tusso	cali	stor	trac	op	mer	sig	sin	ang	ano
con	ac	ag	gam	scar	del	kop	mast	morp	hig

Output: Results obtained with 6th. secondary prefix set as ordered by the default heuristic

Original word	Prefix	Stem
chordamesoderm	chord	amesoderm
chordomesoderm	chord	omesoderm
mercantile	mer	cantile
merge	mer	ge
meringue	mer	ingue
merit	mer	it
meronym	mer	onym
pyracanth	pyr	acanth
sacrilege	sacr	ilege
sacrosanct	sacr	osanct
stroboscope	strob	oscope

Input: 7th. secondary prefix set as ordered by the default heuristic

car	for	bar	ver	in	thing	bur	ove	an	asco
meshug	ter	thin	har	cal	tuss	al	or	cur	tama
obe	budg	coho	ostr	canta	handi	mujah	shall	where	gameto
seismo	roentge	secreta	ar	ser	mor	est	cent	medi	tamar
mar	ran	bors	cor	dis	her	guil	some	am	oxi
ult	cove	fell	ravi	susp	centi	chall	sheat	meshugg	tur
war	ama	el	lan	aya	e'e	aard	alba	aris	azed
bo's	equi	habi	hier	ibid	idio	kara	loll	mave	mulc
nabo	nebb	neph	oosp	piro	roll	teet	vapo	vigo	virt
wron	aspar	baili	baksh	belda	boton	burea	calpa	carca	chitt
chrom	clado	claus	coiff	corte	costu	cring	curra	cusha	devoi
fissi	gibib	goitr	kaffi	khali	kurra	lique	marqu	mebib	megil
nudni	pebib	petab	petti	phala	pillo	pizza	regim	sauer	serap
shill	shitt	silve	sleig	smidg	soign	styra	sulfu	tebib	thrus
tomba	turbo	twili	yobib	zebib	absint	angeli	bolloc	budger	carrag
chemis	chlamy	danseu	fantas	fibrin	kibbut	kolkho	lentis	mackin	pedago
pontif	putref	sapien	satisf	schmal	shella	sidero	sinist	somewh	sprech
suspen	tovari	yashma	yottab	zettab	ankylos	chilias	chutzpa	conidio	golliwo
nomogra	ostraco	pinnati	pleonas	ravigot	sandara	skiagra	sonogra	thingma	cantalou
challeng	chauffeu	gametoph	leitmoti	planocon	spermato	tchotchk	thingama	thinguma	ribonucle
roentgeno	seismogra	xanthophy	highfaluti	tam	crum	hall	hal	ora	wel
long	mish	over	ribo	carac	wor	chur	what	mus	sil
men	cer	gol	por	ber	shel	blan	ava	cach	kibb
kibi	oran	poll	sati	thor	wall	val	mas	sal	cir

cit	blin	lang	kin	vel	ven	bul	gen	int	bil
suc	oce	usu	abys	acol	alky	amon	anne	apac	auro
buck	cair	dall	elas	fond	hicc	homb	hyal	indi	kope
ligh	litt	neig	neve	ninj	oxid	siam	skul	supe	sync
tume	volu	yogh	barbe	branc	carib	champ	fello	kibib	morph
orang	phant	polly	quand	salaa	stoma	trave	velou	wallo	wreat
breath	centim	handic	shallo	strang	mass	arg	star	dan	lin
cas	tus	bill	kind	lent	pila	sand	velo	squir	cel
cen	bor	pil	bas	seri	tera	vers	res	fin	baro
scal	shir	min	aga	hel	ear	kur	coll	larg	mani
phan	resi	breat	char	mes	pur	ten	len	fur	ast
lar	kib	kit	bon	tar	fet	ent	euc	eve	ima
oes	agai	allo	anth	circ	hack	have	holl	madr	meag
napr	negl	nigh	noug	pali	remi	sara	suma	tast	weig
yarm	blint	carre	chang	creas	grand	guill	langu	massi	shtic
terab	whirl	schtic	tamara	tamari	opa	syn	bulg	clim	maca
prim	snar	spur	whor	whir	kal	bir	bis	pas	duc
per	tom	tor	gran	wal	som	cour	dist	leas	plum
sala	ther	bel	gul	nar	cara	as	mit	mac	mat
yar	cul	cus	dir	er	aby	zeb	blit	rang	whel
stran	pun	put	pos	air	ecr	tyr	brus	bunc	comf
dear	galo	geni	glit	gour	hors	intu	kali	knac	legi
peni	pinc	recu	riba	sabo	sens	smar	thum	weal	wild
borsc	borsh	hallu	scall	sprin	tusso	cali	stor	trac	op
sig	sin	ang	ano	con	ac	ag	gam	scar	del
kop	mast	morp	hig	nic	nig	gros	san	ped	ul

Output: Results obtained with 7th. secondary prefix set as ordered by the default heuristic

Original word	Prefix	Stem
pedagog	ped	agog
pedagogue	ped	agogue
pederast	ped	erast

Input: 8th. secondary prefix set as ordered by the default heuristic

car	for	bar	ver	in	thing	bur	ove	an	asco
meshug	ter	thin	har	cal	tuss	al	or	cur	tama
obe	budg	coho	ostr	canta	handi	mujah	shall	where	gameto
seismo	roentge	secreta	ar	ser	mor	est	cent	medi	tamar
mar	ran	bors	cor	dis	her	guil	some	am	oxi
ult	cove	fell	ravi	susp	centi	chall	sheat	meshugg	tur
war	ama	el	lan	aya	e'e	aard	alba	aris	azed
bo's	equi	habi	hier	ibid	idio	kara	loll	mave	mulc
nabo	nebb	neph	oosp	piro	roll	teet	vapo	vigo	virt
wron	aspar	baili	baksh	belda	boton	burea	calpa	carca	chitt
chrom	clado	claus	coiff	corte	costu	cring	curra	cusha	devoi
fissi	gibib	goitr	kaffi	khali	kurra	lique	marqu	mebib	megil
nudni	pebib	petab	petti	phala	pillo	pizza	regim	sauer	serap
shill	shitt	silve	sleig	smidg	soign	styra	sulfu	tebib	thrus
tomba	turbo	twili	yobib	zebib	absint	angeli	bolloc	budger	carrag
chemis	chlamy	danseu	fantas	fibrin	kibbut	kolkho	lentis	mackin	pontif
putref	sapien	satisf	schmal	shella	sidero	sinist	somewh	sprech	suspen
tovari	yashma	yottab	zettab	ankylos	chilias	chutzpa	conidio	golliwo	nomogra
ostraco	pinnati	pleonas	ravigot	sandara	skiagra	sonogra	thingma	cantalou	challeng
chauffeu	gametoph	leitmoti	planocon	spermato	tchotchk	thingama	thinguma	ribonucle	roentgeno
seismogra	xanthophy	highfaluti	tam	crum	hall	hal	ora	wel	long
mish	over	ribo	carac	wor	chur	what	mus	sil	men
cer	gol	por	ber	shel	blan	ava	cach	kibb	kibi
oran	poll	sati	thor	wall	val	mas	sal	cir	cit
blin	lang	kin	vel	ven	bul	gen	int	bil	suc
oce	usu	abys	acol	alky	amon	anne	apac	auro	buck
cair	dall	elas	fond	hicc	homb	hyal	indi	kope	ligh
litt	neig	neve	ninj	oxid	siam	skul	supe	sync	tume
volu	yogh	barbe	branc	carib	champ	fello	kibib	morph	orang
phant	polly	quand	salaa	stoma	trave	velou	wallo	wreat	breath
centim	handic	shallo	strang	mass	arg	star	dan	lin	cas
tus	bill	kind	lent	pila	sand	velo	squir	cel	cen
bor	pil	bas	seri	tera	vers	res	fin	baro	scal
shir	min	aga	hel	ear	kur	coll	larg	mani	phan

resi	breat	char	mes	pur	ten	len	fur	ast	lar
kib	kit	bon	tar	fet	per	ent	euc	eve	ima
oes	agai	allo	anth	circ	hack	have	holl	madr	meag
napr	negl	nigh	noug	pali	remi	sara	suma	tast	weig
yarm	blint	carre	chang	creas	grand	guill	langu	massi	shtic
terab	whirl	schtic	tamara	tamari	opa	syn	bulg	clim	maca
prim	snar	spur	whor	whir	kal	bir	bis	pas	duc
tom	tor	gran	wal	som	cour	dist	leas	plum	sala
ther	bel	gul	nar	cara	as	mit	mac	mat	yar
cul	cus	dir	er	aby	zeb	blit	rang	whel	stran
pun	put	pos	air	ecr	tyr	brus	bunc	comf	dear
galo	geni	glit	gour	hors	intu	kali	knac	legi	peni
pinc	recu	riba	sabo	sens	smar	thum	weal	wild	borsc
borsh	hallu	scall	sprin	tusso	cali	stor	trac	op	sig
sin	ang	ano	con	ac	ag	gam	scar	del	kop
mast	morp	hig	nic	nig	gros	san	ul	ur	tic

Output: No results were obtained with 8th. secondary prefix set as ordered by the default heuristic

Appendix 57

Tertiary concatenation whole word stoplist

acerate	addax	addend	admass	adobe	airscrew
albumin	allice	alphabet	anthem	archive	ascoma
ashram	askant	aspen	automat	axseed	baddie
ballad	bargain	barrack	barrow	bathos	baton
batten	bead	beany	bedlam	begum	bema
benthos	bigos	bingo	binocular	bittie	bobby
bologram	bolograph	booby	boreas	boughten	budget
bugloss	bugology	bulletin	bullion	busby	cabby
cabin	cablegram	campion	canape	cancan	candent
canescent	canfield	canteen	canthus	capsize	capstan
carbide	carbonado	carcase	cargo	carnation	carpet
carrot	cartouch	cartridge	caruncle	cashmere	caterpillar
catsup	centas	chaffinch	champion	chaplet	chewink
chichi	chicken	clamant	claymore	clubable	comedo
coontie	cuppa	cuprite	curfew	curtail	damage
damask	dammar	damson	diesis	dingo	dinkey
discant	docent	dodo	doggo	donkey	donut
dopa	dotage	doubleton	douse	dowager	downward
doyen	dragon	drugget	dryad	earnest	elaterid
eventration	faction	fanfare	fanion	fantan	farad
farrow	farthing	fillagree	finespun	flagon	flexion
fluidram	fluorescein	fluxion	fondant	footslog	formalin
frontlet	furlong	furore	furring	furrow	furuncle
galago	galax	galore	garboil	garbology	gauntlet
gemma	getable	goad	goby	google	goshawk
gosling	gosmore	gossip	gramma	grammar	graphology
gringo	gumma	habitant	halocarbon	hamlet	hammock
hatred	hearken	hellion	hemlock	heroin	hexad
hijab	history	homespun	hotshot	hubby	humin
hummock	indie	indue	ingrate	inion	instar
jambeau	jujube	justice	kentan	kitten	laddie
lambaste	lamprey	landscape	lapin	lappet	laterite
lathi	latten	legend	leghorn	legion	listless
litany	litas	lobby	logion	lotion	lustrate
macaw	madam	madame	mahoe	maidism	maillot
malady	malefactor	malemute	malinger	malope	mandrill
mango	mangold	mangrove	manroot	mansion	manticore
mantiger	mantrap	marabout	margay	margrave	marmite
marrow	marshall	marten	mason	massacre	massage
mastiff	maunder	menace	menage	meteorology	midwife
million	minion	minnow	mission	mixology	moppet
mullion	neoclassic	neocon	neocortex	neoliberal	neonatal
neoplastic	newton	nocent	noma	nomad	nosology
nostrum	notion	novice	nowhere	onion	onward
osprey	outward	overtrump	paddock	padrone	pageant
panache	papain	papaw	papism	pappa	pareve
parget	parrot	parsec	parsnip	parson	partridge
passado	passee	passion	pastern	pastime	pastry
patas	pathos	patten	pause	pawpaw	peasant
penchant	pendragon	pengo	penology	pension	piebald
pierid	pigswill	pillage	pillion	pinion	piperin
piton	plankton	plantar	platform	plumage	plumbago
plumbism	poliosis	poppet	portend	portray	poseuse
postfix	postscript	potable	potage	potion	potlatch
potsherd	potshot	probe	prosthesis	protea	protease
proton	punkey	punnet	puppet	putrid	ragout

rampart	rampion	rapport	ration	redact	redox
reindeer	remittent	rugby	sadism	sagamore	sandhi
sapsago	scandent	scansion	scarlet	schoolgirlish	seascape
season	secant	secpar	secretin	section	seesaw
sergeant	setscrew	shoreward	shylock	sideburn	sidelong
sidereal	siderite	signore	singleton	sirup	sisham
socage	solid	soma	soman	somesthesis	somite
sonnet	soon	soup	soupcon	souse	stallion
stemma	stereophony	stereoscope	stereoscopy	stereotype	strapado
strumpet	summerset	sundry	sunstruck	supraocular	tablespoonful
tanbark	tandoor	tango	tapestry	tappa	tappet
tardive	target	tartar	tartlet	tartrate	tattoo
tautology	teaspoonful	temporise	tenable	tenant	tenno
tenon	tension	theremin	threshold	thumbscrew	thwartwise
tippet	tonsure	topology	topos	tornado	toxicology
traction	tubby	upholster	uppity	upshot	upward
warlock	waterscape	wayward	weirdo	whippet	whitlow
winnow	wolfram	woodscrew	wristlet	writhen	aborad
about	abroach	addax	addend	admass	adobe
adult	aftermath	airdrome	albumen	ampere	aniseed
antelope	anthem	arcane	ardeb	ardour	arete
armoire	arrack	arrow	ascot	ashram	aspen
asphalt	assoil	attune	auriculare	automat	azote
baccarat	bagel	baleen	bandit	bannock	bantam
banting	barbel	barrow	bathe	bayat	beat
beckon	bedlam	benday	benedict	benniseed	benweed
bereave	beroe	besom	betel	bethel	bitok
bittern	bittie	blancmange	blotto	bolete	bollix
bologram	bolograph	bottom	bowel	bowsprit	brandish
bronchoscope	bronchospasm	brothel	bunsen	bunting	burgeon
burrow	bushel	butat	butte	butut	byre
byte	cablegram	cadre	caffre	callathump	camash
camass	camel	camelhair	campong	camwood	canape
candour	canfield	canteen	capote	caput	carat
carburet	carcase	carousel	carpel	carrot	carte
cartel	cartouch	cashmere	casquetel	caterwaul	catsup
caveat	cayuse	centre	certain	chadlock	chaffinch
chapel	charlock	charlotte	chartreuse	chewink	chichipe
chicot	chipper	chiromance	chirrup	chisel	chitchat
chowchow	cismontane	cistern	cityscape	cladding	claim
clamour	clamp	clash	clasp	class	claymore
cleat	clegg	clever	clinch	clink	cloak
clothe	clout	clown	clump	clxv	clxx
cockerel	codex	coiffeuse	colonel	copepod	cornel
cosset	couthie	coxcomb	crabwise	cresson	crowding
cryptanalyst	cudgel	cumquat	cupel	curare	curfew
currycomb	curtail	cutlass	damask	damsel	darkling
darnel	diesel	djinn	dollop	dolmen	dolour
dong	donut	dope	dormie	dossel	dote
douse	doyen	dudeen	duffel	dunnock	duramen
earnest	eastern	eggnog	elbow	encore	endue
ensky	fail	fain	fang	fare	farrow
farthing	fartlek	fastest	fault	fibre	finespun
fluidram	flute	foramen	foredge	format	fornix
forrad	frappe	fringepod	fthm	furlong	furlough
furore	furring	furrow	galax	galere	gallop
galore	gambit	gamete	gamut	gangling	garland
garrote	genre	genteel	germane	gittern	gluten
goat	gong	goniff	goof	google	gook
gore	goshawk	gosmore	gospel	gossip	gout

grippe	grogram	groundsel	gruelling	habitat	hakeem
halogen	haltere	hammock	hareem	hatchel	hatred
hawking	hear	hearse	heart	heartfelt	heel
heft	helm	helot	hemlock	here	hijab
hijack	hippodrome	hire	hobbit	homespun	hostel
hoyden	humane	hummock	jambeau	jujube	kernel
kibe	kibit	kibosh	kickshaw	kidnap	kookie
label	labile	lacrosse	lambast	lambaste	landscape
lariat	latest	latex	lathe	latte	latten
legend	leghorn	levant	level	license	lien
lift	liii	lilac	limen	ling	lintel
lissom	lithe	litre	locomote	locomotor	locoweed
logogram	logograph	logotype	lotte	lungen	lustre
macaw	madam	madame	maglev	magnetograph	magnetosphere
mahoe	maillot	malapropos	malemute	malope	manat
mandrake	mandrill	mangold	mangosteen	mangrove	manticore
marabout	marang	marcel	mare	margrave	marmot
marrow	marshall	marten	martyr	mascot	massacre
masseuse	mastiff	materiel	maxwell	mayhem	megohm
megrim	memsahib	midwife	mien	mildew	milieu
millime	milord	mimeograph	minim	minnow	mire
mitten	moat	model	modem	modern	moderne
mohawk	moil	moire	moloch	molto	momot
month	moolah	mope	more	moreen	mosstone
mote	motel	motmot	moult	mourn	mouse
mung	muscat	muscatel	mushroom	muskat	musquash
mussel	mustache	mustang	naivete	nankeen	napalm
neocortex	neoplasm	newel	newspapering	niblick	nitre
nocent	nook	northern	note	nowhere	nubile
nudibranch	numbat	numen	nutmeg	often	outre
oxen	paddock	padre	palm	palsgrave	panache
panel	pang	pantograph	pantomime	pantothen	papaw
parang	pare	pareve	parrot	parsec	parsnip
partridge	pasang	passee	passel	paste	pastel
pastime	patten	pattern	pause	pavise	paynim
peat	peel	peen	peepul	peeve	peewit
pending	periwig	peruke	pewit	pickaback	pickerel
picot	picul	pilaw	pilot	pinafore	ping
pipe	pipit	pipul	pirogue	pismire	piste
pixel	plaintiff	platen	platyhelminth	plumcot	pointel
pollack	pollen	pollex	pollock	portend	portray
poseuse	probe	prong	proof	proper	protease
proto	pudding	pulpit	pundit	quahog	qualm
quamash	quartern	quasi	radix	ragout	rampart
raphe	rappel	rapport	ratel	realine	rebut
recap	recent	redact	reduce	reel	reeve
refuse	regale	relief	remain	remiss	repair
repast	repent	repine	report	repulse	require
requite	rescue	resect	resent	reside	respire
result	retain	rete	retem	retick	retie
retire	retreat	return	revel	revere	reverse
revile	revolt	rickshaw	ridgel	ringgit	roundel
rowel	rubel	rumen	sachem	sadhe	sagamore
saltire	sardonyx	sateen	scalpel	scarab	scathe
scowling	scrimshaw	secern	secrete	seesaw	sennit
sente	shadblow	shadbush	shaddock	shylock	sicklepod
sideburn	siding	sieve	sift	signore	sincere
sinew	sing	sire	siren	sirup	sisham
skyjack	slattern	soft	solicit	solute	song
sonsie	soon	soothe	sopping	sore	soup

souse	southern	spang	spare	sparrow	spathe
spinel	steppe	stereoscope	stereotype	strophe	swathe
taciturn	takahe	tangram	tarmac	taupe	tautog
teasel	teat	teem	tenting	thousand	threshold
thwartwise	ting	tinsel	tire	tissue	tithe
titre	tittup	together	tope	torte	tote
totem	tout	toward	towel	travelog	tumult
tungsten	umpire	vampire	vandyke	varix	viaduct
vibe	vigilante	viii	virile	visit	vowel
wading	wainscot	wainscotting	warden	webcam	wedel
western	whitlow	whydah	windlass	winnow	withe
witting	wolfram	wombat	writhe		

Appendix 58

Atomic dictionary 1/50 samples prior to stem processing

(with explanations for inclusion)

agin	Spelling variant
amatungulu	Foreign
anywhere	Concatenation component not in WordNet
asp	Atomic
azido	Foreign
bark	Atomic
beg	Atomic
birle	Spelling variant
bliss	Atomic
bond	Irregular quasi-gerund
bow	Atomic
brim	Atomic
bumble	Onomatapoeic
cadastre	Foreign
caracul	Foreign
caw	Onomatapoeic
chanoyu	Foreign
chiliast	Unidentified affix
chutzpah	Foreign
cloche	Foreign
coign	Spelling variant
cosh	Atomic
creak	Onomatapoeic
crump	Onomatapoeic
custom	Irregular Anglo-Norman spelling transformation
danseuse	Foreign
devoice	Missing from Irregular prefix instances
dj	Abbreviation
dreg	Old Norse Gerund
dweeb	U.S. college student slang
emerald	Irregular multilingual derivation
eye	Atomic
feign	Atomic
finesse	Foreign
flight	Irregular quasi-gerund
fondu	Foreign

fringe	Irregular Anglo-Norman spelling transformation
galactagogue	Unidentified affix
geoduck	Foreign
glitz	Back formation
gorge	Atomic
groom	Obscure
gut	Atomic
hang	Atomic
health	Irregular quasi-gerund
high	Atomic
hopple	Spelling variant
hymn	Atomic
inn	Obscure
jihadi	Foreign
kabob	Spelling variant
kibibit	Spelling variant
knockwurst	Foreign
laird	Spelling variant
lcm	Abbreviation
lied	Foreign
logomach	Unidentified affix
luminesce	Unidentified affix
mRNA	Abbreviation
marc	Obscure
meager	Spelling variant
meth	Abbreviation
mm	Abbreviation
moustache	Irregular multilingual derivation
myxomycete	Unidentified affix
neither	Unidentified affix
nog	Obscure
obeah	Foreign
orange	Irregular multilingual derivation
paederast	Spelling variant
peg	Atomic
phlox	Foreign
plank	Irregular Anglo-Norman spelling transformation
pogge	Foreign
pour	Atomic
pseud	Abbreviation
pyelogram	Unidentified affix
quoit	Irregular Anglo-Norman spelling transformation
razmataz	Invention
resume	Erroneous stoplist entry
ritonavir	Unidentified affix
rpm	Abbreviation
sallow	Atomic
scaffold	Irregular Anglo-Norman spelling transformation
sclaff	Obscure
scute	Abbreviation
serif	Irregular multilingual derivation

shelf	Atomic
shote	Obscure
silt	Atomic
slack	Atomic
slur	Atomic
snoot	Back formation
sou	Foreign
spinach	Irregular multilingual derivation
square	Irregular multilingual derivation
steep	Atomic
strake	Obscure
sulfur	Irregular multilingual derivation
swoop	Spelling variant
tandem	Foreign
tench	Atomic
thingamabob	Invention
tight	Atomic
torsk	Obscure
trig	Abbreviation
tun	Atomic
ukase	Foreign
velcro	Abbreviation
vivisect	Unidentified affix
waterborne	Concatenation component not in WordNet
whence	Unhandled inflectional suffix
wind	Atomic
wretch	Irregular quasi-gerund
yack	Onomatapoeic
zag	Foreign

Appendix 59

Stem Dictionary Pruning Algorithm

```
For each stem  in the stem dictionary
{
        the alternative POS for stem is the one (if any) whose corresponding
        POSSpecificLexicalRecord has the most relations of Relation.Type.DERIVATIVE;
        if the stem is not in the main dictionary AND there is an alternative POS AND
        the stem comprises a String of at least 2 characters which is not "ax" then
        {
                for each POSSpecificLexicalRelation of Relation.Type.DERIVATIVE in the
                POSSpecificLexicalRecord associated with the stem
                {
                        the stem derivative is the target of that
                        POSSpecificLexicalRelation;
                        if the stem derivative's POS is the same as the stem's POS then
                        all the POSSourcedLexicalRelations of Relation.Type.ROOT of the
                        POSSpecificLexicalRecord corresponding to the stem derivative as
                        the stem derivative's POS are deleted;
                        a LexicalOmissionException is thrown if the main dictionary does
                        not contain the stem derivative as the stem derivative's POS AND
                        as the alternative POS;
                        if the deleted root relation's target is not the stem AND the
                        stem's prefix list contains the TranslatedPrefix encapsulated in
                        the IrregularPrefixRecord corresponding to the prefix component
                        of the stem derivative then
                        {
                                that TranslatedPrefix is removed from the stem's list of
                                attested prefixes and the DERIVATIVE relation is deleted
                                from the POSSpecificLexicalRecord associated with the
                                stem and all the POSSpecificLexicalRelations of
                                Relation.Type.DERIV of the POSSpecificLexicalRecord
```

```
                                corresponding to the stem derivative as the stem
                                derivative's POS are deleted;
                        }
                }
                if stem has no POSSpecificLexicalRelations left of
                Relation.Type.DERIVATIVE then
                {
                        all LexicalRelations of Relation.Type.ROOT are deleted from the
                        POSSpecificLexicalRecord associated with the stem;
                        if the POSSpecificLexicalRecord associated with the stem still
                        has any Relations which are not of
                        LexicalRelation.SuperType.DERIVATIVE then a
                        DuplicateRelationException is thrown;
                        if the POSSpecificLexicalRecord associated with the stem still
                        has any Relations which are of
                        LexicalRelation.SuperType.DERIVATIVE then
                        {
                                a POSSpecificLexicalRelation of Relation.Type.DERIVATIVE
                                is encoded from the POSSpecificLexicalRecord associated
                                with the stem as the alternative POS to the
                                POSSpecificLexicalRecord associated with the stem as its
                                specified POS;
                                The encoded Relation is written to file "Inter-
                                prefixation relations from stem dictionary pruning.csv";
                                The stem's POS is removed from the entry for the stem in
                                the atomic stem dictionary;
                                if the stem has no other POS, then the entry for the
                                stem is removed from the atomic stem dictionary;
                        }
                        if the POSSpecificLexicalRecord associated with the stem has no
                        Relations left then the stem is removed from the stem
                        dictionary;
                }
        }
}
For each stem in the stem dictionary:
{
        if the stem now has no relations
        {
                the stem is removed from the stem dictionary and the stem's POS from
                the entry for the stem in the atomic stem dictionary.
                If the stem's POS is the only POS given for the stem in the atomic stem
                dictionary, then the entry for the stem is removed from the atomic stem
                dictionary;
        }
}
```

NB The converses of all relations deleted are also deleted.

Appendix 60

Stem meanings

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
acin	N.	sac	N.						nic, ose, us, ar
alumin	N.	aluminium	N.						ate, iferous, ise, ium, ous, um
alveol	N.	cavity	N.						ate, us, ar, ar
apsid	N.	shield	N.					a1, di, syn	dal
arce	N.	arch	N.	bow	N.				ade, ed, us, ella, iform
arch	N.	ruler	N.					ex, matri, mon, patri, olig	
archy	N.	ruler	N.	government	N.			a1, di, matri, mon, patri	
are	N.	dryness	N.						id
aster	N.	star	N.					dis	ral, oid, oid
ax	N.	axe	N.					pole	
ax	N.	axis	N.						il, illa
bacil	N.	bacillus	N.						ary, us, ar, iform
bacter	N.	bacterium	N.						ise, ium, oid, oid
bat	N.	goer	N.					acro	
bat	N.	hitting	N.					con	
bat	N.	bat	N.					mega, micro	
be	N.	life	N.					aero, micro, sapro	
biosis	N.	living	N.	life	N.			aero, ana, anti, cata, crypto, necro, syn	
blast	N.	sprout	N.					ecto, endo, ento, erythro, fibro, hypo, lympho, megalo, meso, mono, myelo, neuro, osteo, melan	ula
blast	N.	blast	N.					counter	
calce	N.	lime	N.	calcium	N.				ed, us, ic, iferous, ite, ium, iform
capit	N.	head	N.						ital, ate, ate, ol
cardium	N.	heart	N.					endo, epi, myo, peri	ia

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
carp	N.	fruit	N.					acro, angio, basidio, endo, epi, exo, meso, mono, peri, pseudo, meri, spor	
cede	V.	go	V.					ad, ante, inter, pre, re, se, super	
cede	V.	yield	V.					con	
ceive	V.	take	V.					con, de, per, re	
cel	N.	cell	N.						ar
cel	N.	small	ADJ.	little	ADJ.			lenti, part	o
cele	N.	hidden	ADJ.					blasto, encephalo, haemato, hemato, hydro, kerato	
cellul	N.	cell	N.						ite, ose, oid, oid
cephaly	N.	head	N.					a1, acro, hydro, macro, mega, megalo, micro, nano, oxy	
cept	N.	taken	ADJ.					con, inter, per, pre	
cess	N.	going	N.					ab, ad, ex, pro, sub	
chlore	N.	chlorine	N.						amine, ide, ine, ite, ella
chrome	ADJ.	colour	N.					bi, mono, poly, tri	
chrome	N.	colour	N.					cyto, fluoro, hemato, mono, poly	
citr	N.	lemon	N.						ic, in, ine, us
claim	V.	shout	V.	cry	V.			ad, counter, de, ex, pro, re	
clase	N.	split	V.					ortho, peri, olig	stic
clave	N.	key	N.	lock	N.			auto, con, en	icle, us
clinal	ADJ.	leaning	ADJ.					ana, anti, cata, iso, syn	
cline	N.	leaning	ADJ.					de, in, mono	
cline	N.	bed	N.						ic
clude	V.	shut	V.	close	V.			con, ex, in, ob, pre, se	
coccus	N.	bacterium	N.					diplo, echino, pneumo, strepto	al
columb	N.	dove	N.	Columbus	N.				ine, ite, ium, o

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
cord	N.	heart	N.					ad, con, dis, re	iform
corn	N.	horn	N.					tri, uni	et
cosm	N.	universe	N.					macro, micro, para	ic, ology
cot	N.	cotyledon	N.					di, mono	
cot	N.	hut	N.	cottage	N.				age, ar
crete	V.	growth	N.					ad, con	
crete	V.	separate	V.					ex, se	
crine	ADJ.	distinguish	V.	separate	V.	judge	V.	apo, ec, endo, exo	
crine	N.	distinction	N.	separation	N.	judgement	N.	endo, exo	
crine	N.	lily	N.						oid, oid
cyte	N.	cell	N.					acantho, astro, blasto, erythro, granul, lympho, macro, megalo, micro, mono, myelo, osteo, thrombo, spher, leuco, melan	ol
derm	N.	skin	N.					blasto, echino, ecto, endo, ento, exo, meso, pachy	
derma	N.	skin	N.					erythro, kerato, scler, xero	
dict	N.	saying	N.					ad, ex, inter, ver	um
dict	V.	say	V.					ad, contra, in, inter, pre	ction
duce	V.	lead	V.					ab, ad, con, de, ex, in, intro, pro, se, trans	
duct	V.	lead	V.					ab, ad, con, de, in	
ennial	ADJ.	yearly	ADJ.					bi, cent, per, tri	
ennial	N.	year	N.					bi, cent, per, tri	
ergy	N.	work	N.					allo, a1, en, syn	
fect	N.	made	ADJ.	done	ADJ.			ad, con, de, ex, pre	
fect	V.	make	V.	done	ADJ.			ad, con, de, ex, in	
fer	N.	bearer	N.	bring	V.			cruci, trans	ry
fer	N.	beast	N.	wild	ADJ.				ral
fer	V.	bring	V.	bear	V.			con, de, dis, in, pre, re, sub, trans	ment

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
ferral	N.	bringing	N.					con, de, re, trans	
ficient	ADJ.	making	N.	do	V.			de, ex, pro, sub	
fit	N.	made	ADJ.					con, pro, bene	
fit	N.	fit	V.					mis, re, retro	
fit	V.	make	V.					pro, bene	
fit	V.	fit	V.					be, re, retro	
flate	V.	blow	V.					con, de, in, re	
flect	V.	bend	V.					de, in, re	ction, exion
flux	N.	flow	N.					con, ex, in, re	
form	ADJ.	shaped	ADJ.					bi, cruci, lenti, multi, uni, vermi	form
form	N.	ant	N.					chloro, fluoro, iodo	ic, ol
form	N.	form	V.					re, uni	ula
form	V.	ant	N.					chloro	
form	V.	form	V.					con, in, per, pre, re, trans, uni	
fract	V.	break	V.					dis, in, re	al, ction, ture
fuge	N.	escape	N.	avoidance	N.	flee	V.	re, vermi, centr, febri, lact	al
fuse	V.	pour	V.					circum, con, de, dis, ex, in, per, sub, trans	
fy	V.	make	V.					cruci, dei, uni	
gamy	N.	marriage	N.	mating	N.			allo, apo, auto, bi, endo, exo, iso, miso, mono, poly	
ge	N.	earth	N.						ology
gen	N.	cause	N.	element	N.			acro, andro, carcino, chromo, cryo, cyano, endo, exo, hydro, immuno, nitro, oxy, patho, pyro, terato, zymo, muta, lact	
gener	N.	kind	N.					con	ral, ic, ic
gest	V.	bring	V.	eat	V.			con, dis, ex, in, sub	
gest	V.	do	V.						ture
gon	N.	angle	N.					dec, epi, hexa, iso, oct, para,	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								penta, peri, poly, tetra, tri	
gram	N.	writing	N.	drawing	N.			aero, ana, angio, arterio, arthro, audio, bi, cardio, crypto, dia, di, echo, encephalo, en, epi, helio, hexa, hist, iso, lipo, mono, myelo, myo, oscillo, penta, pro, radio, spectro, tele, tetra, thermo, tri, holo, ideo, myria, spir, phon, tach	ar
gram	N.	gram	N.					dec, dec, deka, hecto, kilo, micro, milli, nano	
grapher	N.	writer	N.	student	N.			biblio, bio, dem, paleo	
graphy	N.	study	N.	subject	N.	writing	N.	anemo, angio, arterio, arthro, biblio, bio, calli, cardio, crypto, dem, disco, echo, encephalo, epi, hydro, icono, litho, lympho, myelo, ortho, paleo, photo, pyro, radio, tele, thermo, xero	
gress	N.	going	N.					con, ex, in, pro, re	
gress	V.	go	V.					ad, dis, ex, pro, re, retro, trans	
gyny	N.	woman	N.	wife	N.			andro, miso, mono, poly	
hedron	N.	side	N.					dec, hexa, oct, penta, poly, tetra	
herit	V.	inherit	V.					in	able,

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
									age, tor
homin	N.	human being	N.	man	N.				nal, ine, id, oid
hume	N.	earth	N.						ate, ic, in, us, id
hyal	N.	translucent	ADJ.						in, ine, ine, oid, oid
ify	V.	make	V.					acet, aer, electr, ver	
ile	N.	abdomen	N.	entrails	N.				ium
iod	N.	iodine	N.						ide, in, ine, ise
ior	ADJ.	more	ADJ.					exter, infra, inter1, super	
ior	N.	more	ADJ.					exter, infra, inter1, super	
it	N.	going	N.					ad, ex, intro, ob, trans	
it	V.	go	V.					ex, trans	
itis	N.	disease	N.					cephal, entero, gastr, myel, neur, orchi, pneumo, rhin	
ject	N.	thrown	ADJ.					intro, ob, pro, re	
ject	V.	throw	V.					de, ex, in, inter, intro, ob, pro, re	
jure	V.	swear	V.					ab, ad, con, per, NOT_	or
ke	N.	cycle	N.					bi, tri	
kinase	N.	enzyme	N.					entero, strepto, thrombo, ur	
lapse	V.	fall	V.					con, ex, pro, re	
late	V.	bring	V.					dis, ex, re, trans	
late	V.	hide	V.						tent
lateral	ADJ.	side	N.					bi, con, equi, multi, quadr, tri, uni	
latry	N.	worship	N.					anthropo, astro, auto, biblio, demon, helio, icono, idio, mono, pyro, zoo	
lect	N.	gathering	N.					con	
lect	N.	speech	N.	language	N.			dia, idio	
lege	N.	chosen	N.					con, aqu	ate
lege	N.	law	N.						al

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
lepsis	N.	leaving	N.					epan, meta, para, pro, syn	
leptic	N.	leaving	N.					ana, cata, epi, neuro	
lith	N.	stone	N.	rock	N.			entero, hydro, mega, mono, nephro, paleo, xeno, ur, bath, sterco	ic
logue	N.	saying	N.					ana, apo, cata, dia, ec, epi, mono, pro	
logue	N.	speaker	N.					ideo, phil	
logy	N.	study	N.	subject	N.	saying	N.	aero, ana, angio, antho, anthropo, apo, astro, audio, bio, crypto, cyto, derm, ecclesi, eco, eno, entomo, eu, foeto, haemato, hemato, hetero, hist, homo, hydro, immuno, litho, myco, myo, necro, nephro, neuro, osteo, palaeo, paleo, patho, petro, pharmac, phyto, proto, radio, terato, tetra, tri, zoo, zymo	
lude	N.	game	N.	playing	N.			inter, post, pre	o
lude	V.	play	V.					ad, con, de, ex, inter, pre	
lune	N.	moon	N.					apo, peri	ate, ette, ar, ula
lupe	N.	wolf	N.						ine, ine, us
lyse	V.	release	V.					ana, cata, dia, hydro, para	ysis
lysin	N.	liberator	N.	destroyer	N.			cyto, erythro,	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								haemo, hemo, neuro, strepto	
lysis	N.	release	N.	analysis	N.			acantho, auto, bacterio, cyto, electro, haemato, haemo, hemato, hemo, karyo, necro, osteo, pyro, radio, thrombo, zymo	
ma	N.	tumour	N.	growth	N.			acantho, adeno, angio, diplo, fibro, grand, haemato, hemato, lipo, terato	ar, il
mancer	N.	diviner	N.					hydro, litho, necro, pyro, rhabdo	
mancy	N.	divination	N.					hydro, litho, necro, pyro, rhabdo	
mand	V.	order	V.	command	V.	send	V.	con, counter, de	
mant	N.	coat	N.						le, el, illa
mant	N.	prophet	N.						is
medus	N.	jellyfish	N.						ian, an, oid, oid
megaly	N.	enlargement	N.					acro, adeno, cardio, hepato, thyro	
mend	V.	fault	N.					ex, ex	
mend	V.	mind	N.						ntion
mend	V.	hand	N.					con	
mer	N.	part	N.					iso, mono, poly	
mere	N.	part	N.					arthro, blasto, sarco, telo, centr	
metry	N.	measurement	N.					actino, allo, anemo, anthropo, astro, audio, bio, calori, foeto, hydro, iso, micro, photo,	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								psycho, spectro, syn, tele, thermo	
mise	N.	sent	ADJ.	put	V.			de, pre, pro, sur	o
mise	N.	hatred	N.						ology
mise	V.	send	V.	put	V.			de, pre, pro, sur	
mit	V.	send	V.	put	V.			ad, con, ex, inter, intro, man, per, sub, trans	mission
morph	N.	shape	N.	form	N.			allo, ecto, endo, meso, poly, rhizo	ology
mycete	N.	fungus	N.					actino, basidio, blasto, disco, gastro	
mycin	N.	fungus	N.					actino, anti, erythro, myco, strepto	
naut	N.	sailor	N.					aero, aqua, astro, cyber	
nomy	N.	calculation	N.	order	N.	arrangement	N.	a1, anti, astro, auto, eco, gastro	
N.ce	V.	declare	V.	say	V.			ad, de, ex, pro, re	
nym	N.	name	N.					acro, hetero, homo, hyper, pseudo, retro, trop	
oestrous	ADJ.	frenzied	ADJ.	impulsive	ADJ.			a1, di, mono, poly	
oglia	N.	glue	N.					astro, macro, micro, neuro	
oicous	ADJ.	living	ADJ.					hetero, mono, para, poly, syn	
oma	N.	tumour	N.	growth	N.			athero, blasto, carcino, granul, hepato, myelo, myo, neuro, osteo, poly, sarco, xero, zygo	
onym	N.	name	N.					a1, anti, epi, hypo, meta, syn, holo, mer	ous
onymy	N.	name	V.					anti, epi,	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								hypo, syn	
ope	N.	eye	N.					calli, hyper, myo	
opia	N.	eye	N.					a1, hyper, myo, oxy	
opsis	N.	sight	N.	eye	N.			calli, helio, syn	tic
ove	N.	egg	N.						ate, ine, um, iform, oid, oid
pathy	N.	treatment	N.	disease	N.			adeno, allo, angio, arthro, cardio, cryo, encephalo, entero, homeo, hydro, idio, myo, nephro, neuro, osteo, psycho, rhino	
pathy	N.	feeling	N.					anti, en, syn, tele	etic
pe	N.	eye	N.					pyro	
ped	N.	foot	N.					bi, milli, quadr, pinn	dal
pede	N.	foot	N.					milli	ate, icle
pede	N.	child	N.						ology
pel	V.	push	V.					con, dis, ex, in, pro, re	
pend	V.	hang	V.	pay	V.	weigh	V.	ad, de, in	nsion
pene	N.	tail	N.	penis	N.				ial, is
pene	N.	punishment	N.						ology
pete	V.	seek	V.	strive	V.			con	
phage	N.	eater	N.					bacterio, macro, micro, myco	
phagia	N.	eating	N.					aero, a1, dys, necro	
phile	N.	lover	N.					aero, biblio, eno, haemo, hemo, homo, xero	
phile	N.	love	N.						ology
philia	N.	lover	N.					haemo, hemo, necro, para, zoo	
philous	ADJ.	loving	ADJ.					anemo, antho, entomo, phyto	
phone	N.	voice	N.					allo, dia, homo, inter, mega, micro,	ology

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								poly, radio, tele, vibra	
phony	N.	voice	N.					acro, eu, homo, mono, poly, quadr, syn, tele	etic
phore	N.	bearer	N.	bring	V.	carrier	N.	chromo, carpo, spor, gyn	
phyl	N.	leaf	N.					chloro, spor	iform, o
phyll	N.	leaf	N.					cata, chloro, pro, spor	
physeal	ADJ.	growing	ADJ.					apo, dia, epi, hypo	
physis	N.	growth	N.					apo, dia, epi, hypo, meta, para, syn	
phyte	N.	plant	N.					aero, auto, chloro, crypto, epi, hydro, litho, meso, osteo, sapro, xero, zoo, hal, holo, hygro, pterido, spor	
plasia	N.	tissue	N.					ana, a1, cata, dys, hyper, hypo	
plasm	N.	molded	ADJ.	create	V.			cata, cyto, ecto, endo, karyo, nucleo, proto, sarco	
plast	N.	molded	ADJ.	create	V.			amino, chloro, chromo, cyto, proto, pheno	ic
plasty	N.	remold	V.	surgery	N.			ana, angio, arthro, auto, kerato, neuro, rhino	
ple	ADJ.	fold	V.					oct, quadr, sub	
ple	N.	fold	N.					quadr	
ple	V.	fold	V.	bend	V.			quadr, sub	
plegia	N.	stroke	N.	paralysis	N.			di, mono, para, quadr	
plex	ADJ.	woven	ADJ.					con, multi, quadr, tri	
ply	V.	fold	V.					ad, con, in, multi	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
pnea	N.	breath	N.					dys, eu, hyper, hypo, ortho	
pod	N.	foot	N.					actino, amphi, arthro, dec, gastro, hexa, iso, oct, pseudo, rhizo, tetra, tri, myria, ther	
poiesis	N.	making	N.					erythro, haemato, haemo, hemato, hemo, lympho	
port	N.	carry	V.	bring	V.			ex, in, pur, sub, trans	
port	V.	carry	V.	bring	V.			con, de, ex, in, pur, sub, tele, trans	
pose	N.	put	V.					ex, pur, trans	
pose	N.	quantity	N.	dose	N.				ology
pose	V.	put	V.					ad, con, counter, de, dis, ex, in, inter, ob, post, pre, pro, pur, super, sub, trans	
prise	V.	take	V.					ad, con, re, sur	
prive	N.	private	ADJ.						ate, ate, y
proct	N.	rectum	N.	anus	N.			ecto, ento	itis, ology
pteran	N.	winged	ADJ.					di, homo, lepido, neuro	
pute	V.	think	V.					con, de, dis, in	
quan	N.	quantity	N.						ic, ise, um, o
rame	N.	branch	N.						ate, ose, ous, us
rate	V.	rate	V.					be, de, pro, under	ate
rogate	V.	ask	V.	claim	V.	propose	V.	ab, ad, de, inter, sub	ation
rupt	V.	break	V.					dis, ex, inter	ture
sacchar	N.	sugar	N.						ide, in, ine, ose
saur	N.	lizard	N.					allo, megalo, ptero, arch	el
scope	N.	look	V.					angio, arthro, bio, cryo, electro, endo, fluoro, foeto,	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								gastro, icono, kerato, kine, laryngo, micro, ortho, oscillo, peri, pyro, rhino, spectro, tele, chrono, hygro, horo	
scopy	N.	look	V.					arthro, endo, fluoro, foeto, gastro, kerato, micro, radio, rhino, spectro, tele	
scribe	V.	write	V.					ad, circum, de, in, pre, pro, sub, super, trans	
script	N.	written	ADJ.					con, man, pre, re, sub, super, trans	
sect	V.	cut	V.					bi, dis, inter, trans, tri	ction, tor
semble	V.	similar	ADJ.					ad, dis, re	ance
sent	N.	feeling	N.					ad, con, dis	
sert	V.	serve	V.					de	
sert	V.	put	V.	join	V.			ad, ex, in	
serve	V.	save	V.					con, pre, re	
serve	V.	serve	V.					de, sub	
serve	V.	watch	V.					ob	
side	N.	side	N.					a, in, off, under	
sine	N.	sine	N.					arc	
sist	V.	stand	V.	bear	V.			con, de, ex, in, per, sub	
sol	N.	solution	N.					aero, cyto	
sol	N.	sun	N.					para	
sole	N.	comfort	N.					con	
sole	N.	sole	N.					in	
sole	N.	sun	N.						ar
sole	N.	whole	N.						id
sole	N.	alone	ADJ.						o
some	N.	body	N.					acro, auto, chromo, epi, lipo, micro, sarco,	an, ite

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								lyso, centr, prote	
sonate	V.	sound	V.					ad, con, dis, re	
sorb	V.	swallow	V.					ab, ad, de, re	orption
spect	V.	look	V.					ex, in, intro, pro, retro	er
sperm	N.	seed	N.					angio, endo, epi, peri, pteridо, gymn	
spire	V.	breathe	V.					ad, con, ex, in, per, trans	
stat	N.	stationary	ADJ.	stable	ADJ.			bacterio, cryo, haemo, hemo, photo, pyro, thermo, coel	
state	N.	standing	N.					apo, pro	
stitute	V.	set up	V.					con, in, pro, re, sub	
stome	N.	mouth	N.					cyclo, cyto, peri	ate
strate	N.	layer	N.					sub, super	um, us
strict	V.	bind	V.	squeeze	V.	strain	V.	con, dis, re	ture
struct	V.	build	V.					con, de, in, ob	ture
sume	V.	take	V.	eat	V.			ad, con, pre, sub	
tain	V.	hold	V.					ab, ad, con, de, enter, ob, per, sub	
tellur	N.	earth	N.						ian, ic, ide, ium
tend	V.	stretch	V.					ad, con, dis, ex, in, pre, sub	nsion
tene	V.	hold	V.						able, ant, ment, ure, or
tene	V.	hold	V.						able, ant, ment, ure, or
tention	N.	holding	N.					ab, de, ob, re	
test	V.	bear witness	V.					ad, con, de, pro	ator
thelium	N.	establish	V.	stand	V.			endo, epi, meso, peri	
therm	N.	heat	N.					ecto, exo, homeo, homo, iso	
tomy	N.	cutting	N.					amygdal, ana, auto, entero, kerato, litho, myo, nephro, osteo,	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								rhino, rhizo, scler, vaso	
topia	N.	place	N.					dys, ec, sub	ry
tort	V.	twist	V.					con, dis, ex	rsion, ture
tox	N.	poison	N.					de	ic, in, oid
tract	V.	drag	V.	bring	V.			ad, de, dis, ex, pro, sub	ction, tor
tropy	N.	turn	V.					allo, en, ex, iso	
trude	V.	thrust	V.	push	V.			ex, in, ob, pro	
ure	N.	urine	N.						ate, ic, ine, ology
uria	N.	urine	N.					a1, dys, hemat, lymph, poly	
vene	N.	forgiveness	N.						ial
vene	N.	vein	N.						ose, ous, ula
vene	V.	come	V.					contra, con, inter, super	er
vent	V.	come	V.					circum, in, pre, sub	
verse	ADJ.	turned	ADJ.					ad, ab, con, dis, in, per, trans	
verse	N.	turn	N.	side	N.			con, in, ob, uni	o
vert	N.	turned	ADJ.					ad, con, extra, extro, intro, per	
vert	V.	turn	V.					ad, ab, contra, con, dis, ex, intro, in, per, retro, sub	rsion
vious	ADJ.	way	N.					de, ob, per, pre	
vire	N.	virus	N.						ology, us, oid, o
visce	N.	sticky	ADJ.						ose, ous, us, id
vise	V.	seed	N.					ad, de, pre, super, tele	or
visor	N.	see	V.					ad, de, dis, super	
voke	V.	call	V.					con, ex, in, pro	ocation
volve	V.	roll	V.					circum, con, de, ex, in, re	
zoan	ADJ.	animal	ADJ.					ecto, endo, ento, epi, proto	
zoan	N.	animal	N.					actino, antho,	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								ecto, endo, ento, epi, helio, hydro, meta, para, poly, proto	
zoic	ADJ.	living	ADJ.	animal	ADJ.			endo, ento, epi, proto, sapro	
zoon	N.	animal	N.					ecto, ento, epi, proto	
zygous	ADJ.	pair	N.	embryo	N.	gene	N.	a1, di, hetero, homo	
albin	N.	white	ADJ.						nal, nic, ism
alge	N.	seaweed	N.	alga	N.				in, id, oid
algia	N.	pain	N.					cephal, gastr, neur	
ame	N.	ammonia	N.						ide, ine
ammon	N.	ammonia	N.						ium
angin	N.	choking	N.	strangling	N.				ose, ous, na
arsen	N.	arsenic	N.						ate, ic, ide
aur	N.	earth	N.						icle, iform
aur	N.	gold	N.						iferous
aw	N.	awe	N.						ed, ful, less
bare	N.	barium	N.						ic, ite
bitumin	N.	bitumen	N.						ise, ous, oid
bola	N.	throw	N.	trajectory	N.			hyper, meta, para	
bole	N.	throw	N.	trajectory	N.			amphi, hyper	o
bolise	V.	throw	V.					cata, dia, meta	
botul	N.	sausage	N.						in, ism, iform
bove	N.	cattle	N.						ine, ine, id
brach	N.	arm	N.					amphi, di	ium
bronch	N.	windpipe	N.						ial, us, o
bure	N.	jug	N.						et, ette, in
caine	N.	cocaine	N.					benzo, pro, tetra	
capnia	N.	smoke	N.					a1, hyper, hypo	
capt	V.	take	V.	catch	V.				tion, tor, ture
cardia	N.	heart	N.					dextro, mega, megalo	
ceed	V.	go	V.					ex, pro, sub	
cephalus	N.	head	N.					hydro, lepto, micro	
ceps	N.	head	N.					bi, quadr, tri	
cept	V.	take	V.	catch	V.			ad, ex,	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
								inter	
ceptor	N.	taker	N.	catcher	N.			entero, pre, re	
ceram	N.	pottery	N.						ic, ic, ist
cern	V.	sift	V.					con, dis, se	
cess	V.	going	N.					ad, pre, pro	
cessor	N.	go	V.					inter, pro, sub	
chaete	N.	hair	N.					poly, olig, spir	
chezia	N.	defecation	N.					dys, haemato, hemato	
chromia	N.	colour	N.					a1, di, mono	
chrone	N.	time	N.					iso	icle, ology
cide	N.	killing	N.					matri, patri, vermi	
cilie	N.	eyelash	N.						ary, ate, ate
cise	V.	cutting	N.					circum, ex, in	
cite	V.	rouse	V.	summon	V.			ex, in, re	
cline	V.	lean	V.					de, in, re	
clivity	N.	slope	N.					ad, de, pro	
coele	N.	cavity	N.					blasto, haemato, hemato	
cogn	N.	know	V.						ise
come	N.	come	V.					in	
come	N.	hair	N.						et
coron	N.	crown	N.						et, na, illa
crat	N.	ruler	N.					auto, dem, techn	
crement	N.	growth	N.					de, in	
crement	N.	sift	V.					ex	
cumbent	ADJ.	lie down	V.					ad, de, pro	
cune	N.	wedge	N.						ate, us, iform
cur	V.	run	V.					con, in, ob	
cuss	V.	shake	V.					con, dis, per	
dactyl	ADJ.	finger	N.					hetero, poly, zygo	
dactyly	N.	finger	N.					a1, hyper, syn	
demic	ADJ.	people	N.					ec, epi, pan	
dicate	V.	proclaim	V.					ab, de, in	
dign	ADJ.	worthy	ADJ.					con	ify, ity
dolent	ADJ.	suffering	ADJ.					con, in, re	
done	V.	give	V.					con	ee, or
dontia	N.	tooth	N.					endo, exo, ortho	
dontist	N.	dentist	N.					endo, exo, ortho	
dow	V.	give	V.					en	er, er
dox	ADJ.	teaching	N.					hetero, ortho	y

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
dress	V.	straighten	V.					ad, re	
dress	V.	dress	V.					under	
drome	N.	running	N.					aero, pro, syn	
dromous	ADJ.	running	ADJ.					ana, cata, dia	
duct	N.	lead	V.					ad, con, pro	
duct	V.	lead	V.					ab, de, in	
dural	ADJ.	hard	ADJ.					epi, extra, sub	
dure	N.	hard	ADJ.						ess, um
emia	N.	blood	N.					a1, hydr, hyper	
eresis	N.	take	V.					dia, dia, syn	
ethn	N.	race	N.						ic, nic, ology
fasce	N.	bundle	N.						s, icle, ism
fece	N.	stool	N.	excrement	N.				al, s, ula
femin	N.	woman	N.						ine, ine, ise
fine	V.	delimit	V.					con, de	
fine	V.	purify	V.					re	
fine	ADJ.	fine	ADJ.					hyper, super	
fine	ADJ.	bounded	ADJ.	limited	ADJ.				itude
flict	V.	strike	V.					ad, con, in	
flore	N.	flower	N.						et, id
fung	N.	fungus	N.						ous, us, oid
gee	N.	earth	N.					apo, con, peri	
gnosis	N.	knowledge	N.					dia, pro, tele	
gnostic	ADJ.	knowing	ADJ.					dia, pro, tele	
gone	N.	born	ADJ.	offspring	N.	seed	N.	epi, iso, peri	
habit	V.	live	V.					co, in	tant
hale	N.	salt	N.						ide, ite, o
hale	V.	breathe	V.					ex, in	
helion	N.	sun	N.					apo, para, peri	
here	V.	sticky	ADJ.					ad, co, in	
hibit	V.	have	V.	hold	V.			ex, in, pro	
hile	N.	little	ADJ.	small	ADJ.				um, us, ar
hume	V.	earth	N.					ex, in	
ient	ADJ.	go	V.					ab, ad, ambi	
jacent	ADJ.	lie down	V.					ad, sub, super	
jove	N.	Jupiter	N.					apo, peri	ial
junct	ADJ.	joined	ADJ.					ad, con, dis	
karyote	N.	kernel	N.					a1, eu, pro	
kete	N.	acetone	N.						amine, one, ose
labe	N.	take	V.					astro	
labe	N.	lip	N.						ium

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
labe	N.	rag	N.						el
lanthan	N.	hide	V.						ide, um, oid
lapse	N.	fall	V.					con, pro, re	
lect	V.	gather	V.					con	
lect	V.	read	V.						tor, ture
lectic	N.	reading	N.					cata, dys	
lectic	N.	gathering	N.					ec	
lectic	ADJ.	read	V.					cata, dys	
lectic	ADJ.	gather	V.					ec	
lege	V.	choose	V.						acy
lemma	N.	take	V.					di	
lemma	N.	membrane	N.					neuro, sarco	
lepsy	N.	leaving	N.					cata, epi, nympho	
leptic	ADJ.	leave	V.					ana, cata, epi	
leve	V.	raise	V.						ee, er, er
lign	N.	wood	N.						in, ite, um
log	N.	saying	N.	account	N.	ratio	N.	ana, dia, epi	
logist	N.	speaker	N.					electro, mono	istic
lunary	ADJ.	lunar	ADJ.					sub, super, trans	
mage	N.	priest	N.	sorcerer	N.				ic, ic, us
magn	N.	great	ADJ.	large	ADJ.	big	ADJ.		ate, um
magn	N.	lodestone	N.						et
mand	N.	order	N.	command	N.			con, counter, de	
mastigote	N.	whip	N.					hyper, poly, zoo	
mede	N.	middle	N.						ian, ium
mede	N.	healer	N.						ic
ment	N.	mind	N.					con	ntal, um
merous	ADJ.	part	N.					allo, penta, tetra	
metric	ADJ.	measure	V.					dia, para, tetra	
minent	ADJ.	stand out	V.	jut out	V.	protrude	V.	ex, in, pro	
mnemon	N.	memory	N.	reminder	N.				ic, nic, ist
mode	N.	manner	N.	fashion	N.			con	ish, el
mongol	N.	Mongol	N.						ism, oid, oid
mony	N.	state	N.	condition	N.			acri, matri, patri	
mora	N.	snout	N.	muzzle	N.				ine
mora	N.	custom	N.						le
mote	V.	move	V.					de, ex, pro	
muce	N.	mucus	N.						iferous, in, us
mural	ADJ.	wall	N.					extra, inter, intra	
mute	V.	change	V.					con, per, trans	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
nate	ADJ.	born	ADJ.					ad, ex	
nautic	ADJ.	sailor	N.					aero, astro	ical
nomial	N.	calculation	N.	order	N.	arrangement	N.	bi, multi, poly	
nomial	ADJ.	calculate	V.	ordered	ADJ.	arranged	ADJ.	bi, multi, poly	
nove	N.	new	ADJ.						ice, el, ella
ode	N.	way	N.	road	N.			ana, di, tetr	
ody	N.	song	N.					mono, para	
ody	N.	hate	N.						ious
oecious	ADJ.	living	ADJ.					hetero, mono, syn	
omatous	ADJ.	swollen	ADJ.					carcino, granul, neuro	
on	ADJ.	one	ADJ.						ly, ly
orchidism	N.	testicle	N.					a1, crypto, mono	
orchism	N.	testicle	N.					a1, crypto, mono	
ord	V.	rank	N.	series	N.				er, er
ord	V.	filthy	ADJ.						ure
ose	N.	carbohydrate	N.	sugar	N.			dextro, poly, tetr	
pal	V.	pale	ADJ.					ad	or
pand	V.	spread	V.					ex	
pane	N.	cloth	N.					counter	el
pane	N.	fat	N.					pro	
pape	N.	pope	N.						pal, ism
pape	N.	breast	N.	nipple	N.				illa
pape	V.	pope	N.						acy
pape	V.	papyrus	N.						er, er
pede	V.	foot	N.					in	al
pede	V.	child	N.	pupil	N.				ant
pedia	N.	child	N.	teaching	N.			cyclo, hypno, miso	
penia	N.	deficiency	N.					cyto, lympho, thrombo	
pept	N.	cooked	ADJ.						ide, ise, one
phagous	ADJ.	eat	V.					antho, sapro, zoo	
phagy	N.	eating	N.					anthropo, myco, necro	
phasia	N.	speech	N.					a1, cata, dys	
phora	N.	bear	V.	bringing	N.	carry	V.	ana, epan, epi	
phoresis	N.	bear	V.	bringing	N.	carry	V.	cata, dia, electro	
physial	ADJ.	growing	ADJ.					dia, epi, hypo	
plete	V.	fill	V.					con, de, re	
plex	N.	woven	ADJ.					con, multi	us
plicity	N.	fold	N.					con, multi, tri	
plode	V.	clap	V.					ex, in	sion

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
ploid	ADJ.	shaped	ADJ.	chromosome	N.			mono, poly, tri	
plore	V.	cry	V.					de, in	
pode	N.	foot	N.					anti, mega	ium
polis	N.	city	N.	state	N.			acro, megalo, necro	
port	V.	carry	V.					con, de, tele	
pos	N.	foot	N.					tri	
posit	V.	put	V.					de, ex, re	
pository	N.	put	V.					de, re, sub	
pot	N.	put	V.					de, inter	o
pote	V.	drink	V.						able
pote	V.	pot	N.						age
pound	V.	put	V.					ex, in, pro	
prove	V.	try	V.	test	V.			ad, en, re	
pteron	N.	wing	N.					di, lepido, neuro	
pugn	V.	fight	V.					in, ob, re	
punct	N.	point	N.	dot	N.				ual, uate, um
pus	N.	foot	N.					oct, rhizo	
que	N.	asking	N.	seeking	N.	getting	N.		ery
quest	N.	asking	N.	seeking	N.	getting	N.	con, in	
quire	V.	ask	V.	seek	V.	get	V.	ad, en, in	
rach	N.	spine	N.						is, itis
rect	N.	straight	ADJ.						um, us
rect	N.	right	ADJ.						o
rect	ADJ.	right	ADJ.	straight	ADJ.				ify, itude
ren	N.	kidney	N.						nal
ren	N.	curdling	N.						et, in
reve	N.	dream	N.						ery
reve	N.	rebel	N.						el
rheumat	N.	stream	N.						ism, ology, oid
rive	V.	shore	N.	river	N.			ad, de	er
rode	V.	gnaw	V.					con, ex	dent
sanct	ADJ.	holy	ADJ.						ify, itude, ity
scand	V.	trap	V.	tempt	V.				al
scand	V.	climb	V.						ndent
scand	V.	scan	V.						nsion
scend	V.	climb	V.					ad, de, trans	
scient	ADJ.	knowing	ADJ.					omni, pre	nce
scopic	ADJ.	look	V.					acro, macro, mega	
secutor	N.	follower	N.					ex, per, pro	
semin	N.	seed	N.						nal, iferous, ar
sent	V.	feel	V.					ad, con, dis	
sert	N.	joined	ADJ.	put	V.			de, in	

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
sert	N.	serve	V.					dis	
serve	V.	serve	V.					de, sub	
serve	V.	watch	V.					ob	
sess	V.	sit	V.					ad, ob	ssion
shore	ADV.	shore	N.					a, in, off	
side	N.	side	N.					a, under	
sile	N.	barn	N.						o
sile	V.	barn	N.					en	age
solute	ADJ.	free	ADJ.	separated	ADJ.	loosen	V.	ab, dis, re	
solve	V.	free	V.	separate	V.	loosen	V.	ab, dis, re	
somy	N.	chromosome	N.					mono, poly, tri	
son	N.	song	N.					grand	
son	N.	song	N.	sound	N.			uni	et
spect	N.	look	N.					ad, pro, retro	
sperse	V.	scatter	V.					ad, dis, inter	
spond	V.	answer	V.					de, re	
stal	V.	stand out	V.	stable	N.			in	llion, ll
stasia	N.	standing	N.					a1, haemo, hemo	
stasy	N.	standing	N.					apo, ec, iso	
stere	N.	solid	N.	cholesterol	N.				oid, ol, o
stitute	N.	set up	V.					in, pro, sub	
stole	N.	sent	ADJ.	put	V.			dia, syn	
stylar	ADJ.	columnar	ADJ.					amphi, a1, peri	
style	N.	column	N.					cyclo, peri, sarco	
suade	V.	urge	V.					dis, per	sion
sult	V.	jump	V.	leap	V.			con, ex, in	
sure	V.	secure	V.	safe	ADJ.			ad, en, in	
tarant	N.	tarantula	N.						ism, ella, ula
taxy	N.	arrangement	N.					a1, epi, hetero	
tene	N.	band	N.	ribbon	N.			diplo, lepto	
tene	N.	held	ADJ.						et
terr	V.	earth	N.						ace
terr	V.	frighten	V.					ible, or	
test	V.	bear witness	V.					ad, de	ator
thal	N.	sprout	N.						ium, us, oid
thene	N.	palm	N.						ar, ar
toment	N.	down	N.	stuffing	N.				ose, ous, um
ton	N.	ton	N.					kilo, mega	
tope	N.	place	N.					epi, iso	ology
trope	N.	turn	N.					allo, helio	ism
trophy	N.	nourishment	N.					a1, dys, hyper	
tropous	ADJ.	turn	V.					amphi, ana, ortho	
turb	N.	eddy	N.						ine, id

Stem		Meanings							
Form	**POS**	**Word**	**POS**	**Word**	**POS**	**Word**	**POS**	**Prefixes**	**Suffixes**
uresis	N.	urine	N.					a1, dia, en	
vade	V.	go	V.					ex, in, per	
vail	V.	worth	ADJ.					ad, counter, pre	
valve	N.	shutter	N.	door	N.			bi, uni	ula
vare	N.	variety	N.						iform
vect	V.	convey	V.	carry	V.			ad, con	tor
vele	N.	sailor	N.	curtain	N.				um, ar, ar
venge	V.	avenge	V.					a, re	ance
vent	N.	coming	N.					ad, con, ex	
V.	N.	word	N.	V.	N.			ad, pro	al
vey	V.	travel	V.					con	
vey	V.	see	V.					pur, sur	
veyor	N.	traveller	N.					con	
veyor	N.	see	V.					pur, sur	
vict	V.	win	V.	conquer	V.	overcome	V.	con, ex	tor
vince	V.	win	V.	conquer	V.	overcome	V.	con, ex	ible
vulcan	N.	fire	N.						ise, ite, ology
xanth	N.	yellow	ADJ.						ate, ine, ous
xyle	N.	wood	N.						ne, ose, ol

Appendix 61

Encoding of relations between stems and their components

Parameters

Parameter	Type
`analysedAffixationComponents`	`Map<POSTaggedStem, List<Morpheme>>`
`lexicalRestorationStoplist`	`Set<POSTaggedMorpheme>`
`includeInterpreted`	`Boolean`
`lexicalRestorationsFile`	`OutputFile`

Parameter `includeInterpreted` specifies whether POSTaggedStems which have been interpreted are to be included in the analysis.

```
For each entry in analysedAffixationComponents:
{
        POSTaggedStem derivative is the key and List<Morpheme> components is the value,
        If includeInterpreted is true or if derivative has not been interpreted
        {
                For each Morpheme component in components:
                {
                        if component is a POSTaggedStem
                        {
                                if component is in the main dictionary as its specified
                                POS and is not in lexicalRestorationStoplist and is not
                                monosyllabic
                                {
                                        A POSSpecificLexicalRelation of
                                        Relation.Type.DERIVATIVE and
                                        LexicalRelation.SuperType.DERIVATIVE is encoded
                                        from the POSSpecificLexicalRecord corresponding
                                        to component as a POSTaggedWord to derivative as
                                        a POSTaggedStem and its converse
                                        POSSpecificLexicalRelation from the
```

```
                        POSSpecificLexicalRecord corresponding to
                        derivative to component.
                        derivative and component are written to
                        lexicalRestorationsFile
                }
                Otherwise
                {
                        A POSSpecificLexicalRelation of
                        Relation.Type.ROOT and
                        LexicalRelation.SuperType.ROOT is encoded from
                        the POSSpecificLexicalRecord corresponding to
                        derivative to component as a POSTaggedStem and
                        its converse POSSpecificLexicalRelation from the
                        POSSpecificLexicalRecord corresponding to
                        component to derivative.
                        the stem dictionary and atomic stem dictionary
                        are updated with component , its affix list and
                        its POS
                }
        }
        Otherwise if component is a TranslatedPrefix:
        {
                for each of its meanings:
                {
                        A translating POSSpecificLexicalRelation of
                        Relation.Type.ROOT and
                        LexicalRelation.SuperType.ROOT is encoded from
                        the POSSpecificLexicalRecord corresponding to
                        derivative as a stem to meaning and its converse
                        POSSpecificLexicalRelation of
                        Relation.Type.DERIVATIVE and
                        LexicalRelation.SuperType.DERIVATIVE from the
                        POSSpecificLexicalRecord corresponding to meaning
                        to derivative. If one or other of the relation to
                        be encoded and its converse (but not both) is
                        already encoded or if the same Relation is
                        already encoded as a different subclass of
                        LexicalRelation then a POSTargetedLexicalRelation
                        is encoded from the GeneralLexicalRecord
                        corresponding to derivative with converse
                        POSSourcedLexicalRelation. If this latter
                        relation or its converse (but not both) is
                        already encoded or if the latter Relation is
                        already encoded as a different subclass of
                        LexicalRelation then meaning is converted to
                        uppercase and another attempt is made to encode a
                        POSSpecificLexicalRelation and converse
                        POSSpecificLexicalRelation. If this latter
                        relation or its converse (but not both) is
                        already encoded or if the latter Relation is
                        already encoded as a different subclass of
                        LexicalRelation then a POSTargetedLexicalRelation
                        is encoded from the GeneralLexicalRecord
                        corresponding to derivative with converse
                        POSSourcedLexicalRelation.
                }
        }
        Otherwise if component is a POSTaggedSuffixation:
        {
                 If component is in the main dictionary as its specified
                POS and is not in lexicalRestorationStoplist and does
                not represent a monosyllabic word:
                {
                        A POSSpecificLexicalRelation of the converse type
                        of Relation.Type stored in component as a
                        POSTaggedSuffixation is encoded from the
                        POSSpecificLexicalRecord corresponding to
                        component as a POSTaggedSuffixation as a
                        POSTaggedWord to derivative as a POSTaggedStem
                        and its converse POSSpecificLexicalRelation from
                        the POSSpecificLexicalRecord corresponding to
                        derivative as a POSTaggedStem to component.
                        and derivative and its POS, followed by component
                        and its POS are written to
                        lexicalRestorationsFile.
                }
```

```
                                Otherwise, provided that component as a
                               POSTaggedSuffixation represents some word form:
                               {
                                       the POSTaggedStem representation of component as
                                       a POSTaggedSuffixation is added to the stem
                                       dictionary and its wordform is added to the
                                       atomic stem dictionary (if not already present)
                                       and its POS is added to the POSes mapped to in
                                       the atomic stem dictionary  by its wordform.
                                       and a POSSpecificLexicalRelation of the type
                                       stored as component as a POSTaggedSuffixation's
                                       Relation.Type is encoded from the
                                       POSSpecificLexicalRecord corresponding to
                                       derivative as a POSTaggedStem to component as a
                                       POSTaggedSuffixation and its converse
                                       POSSpecificLexicalRelation from the
                                       POSSpecificLexicalRecord corresponding to
                                       component as a POSTaggedSuffixation to derivative
                                       as a POSTaggedStem.
                               }
                        }
                }
        }
}
```

Appendix 62

Generic disambiguation Algorithm One by One

```
reader = new GoldStandardReader();
window = new DisambiguationContextWindow();
reset paradox count to 0;
output = new List<DisambiguationOutputWord>();
cntr = 0;
while (cntr < window.size())
{
        nextWindowOccupant = reader.getNextOccupant();
        window.advance(nextWindowOccupant);
        cntr++;
}
while (nextWindowOccupant != null)
{
        nextWindowOccupant = reader.getNextOccupant();
        DisambiguationOutputWord latestOutput = window.advance(nextWindowOccupant);
        output.add(latestOutput);
}
cntr = 0;
while (cntr < window.size())
{
        DisambiguationOutputWord latestOutput = window.advance(null,);
        output.add(latestOutput);
        cntr++;
}
return output;

DisambiguationOutputWord
DisambiguationContextWindow.advance(DisambiguationWindowOccupant newOccupant)
{
        windowOccupants.add(newOccupant);
        DisambiguationWindowOccupant windowLeaver = windowOccupants.remove();
        DisambiguationWindowOccupant target = windowOccupants.get(targetIndex);
        if (target.disambiguable()
        {[22]
                bestWordSenses = disambiguate(target, senseMatchMeasure, false);
                if (bestWordSenses is null)
                {
                        bestWordSenses = disambiguate(target, senseMatchMeasure, true);
                }
                if (bestWordSenses is null)
                {
```

[22] The `List<SenseCombination>` is created here for the B&P and Nearest Neighbours algorithms (§§6.3.6.2.3, 6.3.6.3)

```
                        bestWordSenses = disambiguate(target, glossOverlapMeasure,
                        false);
                }
                if (bestWordSenses is null)
                {
                        bestWordSenses = disambiguate(target, glossOverlapMeasure,
                        true);
                }
                if (bestWordSenses is null)
                {
                        disambiguateByFreqency(target);
                        target.recordDefault();
                        return;
                }
                for (each currentBestSense in bestWordSenses)
                {
                        if (currentBestSense is not null)
                        {
                                if (currentBestSense is in target position)
                                {
                                        if (target.bestSense is null)
                                        {
                                                target.bestSense = currentBestSense;
                                        }
                                        else if (target.bestSense is not
                                        currentBestSense)
                                        {
                                                target.bestSense = currentBestSense;
                                                target.recordParadox();
                                                increment paradox count;
                                        }
                                }
                                else
                                {
                                        otherOccupant = DisambiguationWindowOccupant in
                                        position corresponding to
                                        currentBestSense
                                        if (otherOccupant.bestSense is null)
                                        {
                                                otherOccupant.bestSense =
                                                currentBestSense;
                                        }
                                        else
                                        {
                                                if (otherOccupant.bestSense
                                                is not currentBestSense)
                                                {
                                                        otherOccupant.recordParadox();
                                                        increment paradox count;
                                                }
                                        }
                                }
                        }
                }
        }
        return new DisambiguationOutputWord(windowLeaver.word, windowLeaver.bestSense,
        windowLeaver.paradoxical, windowLeaver.defaulted, windowLeaver.disambiguable);
}

List<WordSense>  DisambiguationContextWindow.disambiguate(DisambiguationWindowOccupant
target, RelatednessMeasure thisMeasure, Boolean heavy)[23]
{
        bestSenses = new List<WordSense>();
        bestScore = 0;
        for (each occupant in windowOccupants)
        {
                if (occupant is not target)
                {
                        WordSense[] currentBestSenses = target.disambiguate
                        (occupant, thisMeasure, heavy, morphologicalAwareness);
                        if (currentBestSenses is null)
                        {
                                bestSenses.add(null);
                        }
```

[23] B&P and Nearest Neighbours algorithms as described (§§6.3.6.2.3, 6.3.6.3) replace this method.

```
                        else
                        {
                                score = target.currentScore();
                                if (score is equal to bestScore)
                                {
                                        bestTargetSense = null;
                                        bestSenses.add(null);
                                }
                                else
                                {
                                        if (score > bestScore)
                                        {
                                                bestScore = score;
                                                bestTargetSense =
                                                currentBestSenses[local];
                                                bestSenses.add
                                                (currentBestSenses[remote]);
                                        }
                                        else
                                        {
                                                bestSenses.add(null);
                                        }
                                }
                        }
                }
                else
                {
                        bestSenses.add(null);
                }
        }
        if (bestTargetSense == null)
        {
                return null;
        }
        bestSenses.set(targetIndex, bestTargetSense);
        return bestSenses;
}

WordSense[]      DisambiguationWindowOccupant.disambiguate(DisambiguationWindowOccupant
other,          RelatednessMeasure          thisMeasure,          Boolean          heavy,
Disambiguator.MorphologicalAwareness morphologicalAwareness)
{
        bestWordSenses = new WordSense[2];
        bestScore = 0;

        for (each WordSense thisWordSense in this.possibleSenses)
        {
                for (each WordSense otherWordSense in other.possibleSenses)
                {[24]
                        switch (morphologicalAwareness)
                        {
                                case LEXICAL:
                                {
                                        theseSynsets = this.lexicalRelativesLists.get
                                        (thisWordSense).synsets();
                                        otherSynsets = other.lexicalRelativesLists.get
                                        (otherWordSense).synsets();
                                        break;
                                }
                                case SEMANTIC:
                                {
                                        theseSynsets = this.semanticRelativesLists.get
                                        (thisWordSense).synsets();
                                        otherSynsets = other.semanticRelativesLists.get
                                        (otherWordSense).synsets();
                                        break;
                                }
                                case MORPHO_SEMANTIC:
                                {
                                        theseSynsets = this.semanticRelativesLists.get
                                        (thisWordSense).synsets();
                                        otherSynsets = other.semanticRelativesLists.get
```

[24] The contents of this loop are also executed by the B&P algorithm (§§6.3.6.2.3) when calculating the score of a `SenseCombination`.

```
                                        (otherWordSense).synsets();
                                        theseSynsets.addAll
                                        (this.lexicalRelativesLists.get
                                        (thisWordSense).synsets());
                                        otherSynsets.addAll
                                        (other.lexicalRelativesLists.get
                                        (otherWordSense).synsets());
                                        break;
                                }
                        }
                        if (heavy)
                        {
                                score = thisMeasure.measure(theseSynsets, otherSynsets);
                        }
                        else
                        {
                                thisSynset = wordnet.fetchSynset(thisWordSense);
                                otherSynset = wordnet.fetchSynset(otherWordSense);
                                score = thisMeasure.measure(thisSynset, otherSynset,
                                theseSynsets, otherSynsets);
                        }
                        if (score is equal to bestScore)
                        {
                                bestWordSenses[local] = null;
                                bestWordSenses[remote] = null;
                        }
                        else if (score > bestScore)
                        {
                                bestScore = score;
                                bestWordSenses[local] = thisWordSense;
                                bestWordSenses[remote] = otherWordSense;
                        }
                }
        }
        currentScore = bestScore;
        if (bestWordSenses[local] == null)
        {
                return null;
        }
        return bestWordSenses;
}
```

Appendix 63

Disambiguation results

Key

Ww. size	Window size
MORPH. AWARENESS	MORPHOLOGICAL AWARENESS (tables 53-54)
LEX. RELTY.	LEXICAL RELATIVITY (tables 53-54)
W	disambiguable words
f	failures (no disambiguation result)
d	defaults (disambiguated by frequency; excluding failures)
p	paradoxes (§6.3.6.1.1)
C_{-d}	correct non-defaults
C_{+d}	correct defaults
R	Recall
P	Precision
C_v	Coverage

B&P Algorithm

Ww. size	MORPH. AWARENESS	LEX. RELTY.	W	f	d	p	C_{-d}	C_{+d}	R	P	C_v
3	SEMANTIC	NON- LEXICAL	2421	305	1326	139	417	822	17.22%	52.78%	32.63%
	LEXICAL	SYNONYMOUS	2421	296	1131	126	339	710	14.00%	34.10%	41.06%
		SEMANTICALLY-RELATED	2421	234	690	209	743	417	30.69%	49.63%	61.83%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	249	775	211	621	478	25.65%	44.45%	57.70%
		SEMANTICALLY-RELATED	2421	231	670	204	758	401	31.31%	49.87%	62.78%
5	SEMANTIC	NON- LEXICAL	2421	319	1630	234	251	992	10.37%	53.18%	19.50%
	LEXICAL	SYNONYMOUS	2421	298	1398	290	236	869	9.75%	32.55%	29.95%
		SEMANTICALLY-RELATED	2421	218	914	420	643	555	26.56%	49.88%	53.24%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	230	1034	462	506	638	20.90%	43.73%	47.79%
		SEMANTICALLY-RELATED	2421	209	884	421	667	536	27.55%	50.23%	54.85%
	Baseline		**2421**	**427**	**1994**	**0**	**0**	**1206**	**49.81%**	**60.48%**	**82.36%**

Nearest Neighbours Algorithm

Ww. size	MORPH. AWARENESS	LEX. RELTY.	W	f	d	p	C_{-d}	C_{+d}	R	P	C_v
3	SEMANTIC	NON- LEXICAL	2421	305	1325	139	418	821	17.27%	52.84%	32.67%
	LEXICAL	SYNONYMOUS	2421	296	1131	126	339	710	14.00%	34.10%	41.06%
		SEMANTICALLY-RELATED	2421	234	690	209	743	417	30.69%	49.63%	61.83%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	249	775	211	621	478	25.65%	44.45%	57.70%
		SEMANTICALLY-RELATED	2421	231	670	204	758	401	31.31%	49.87%	62.78%
5	SEMANTIC	NON- LEXICAL	2421	275	1354	254	417	820	17.22%	52.65%	32.71%
	LEXICAL	SYNONYMOUS	2421	272	1163	257	349	726	14.42%	35.40%	40.73%
		SEMANTICALLY-RELATED	2421	222	706	364	747	425	30.86%	50.03%	61.67%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	226	787	407	621	480	25.65%	44.11%	58.16%
		SEMANTICALLY-RELATED	2421	216	679	361	778	405	32.14%	50.98%	63.03%
7	SEMANTIC	NON- LEXICAL	2421	273	1377	285	407	845	16.81%	52.79%	31.85%
	LEXICAL	SYNONYMOUS	2421	251	1162	329	361	731	14.91%	35.81%	41.64%
		SEMANTICALLY-RELATED	2421	186	730	482	776	443	32.05%	51.56%	62.16%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	201	821	534	610	510	25.20%	43.60%	57.79%
		SEMANTICALLY-RELATED	2421	185	715	473	785	430	32.42%	51.61%	62.83%
11	SEMANTIC	NON- LEXICAL	2421	272	1383	302	413	859	17.06%	53.92%	31.64%
	LEXICAL	SYNONYMOUS	2421	241	1179	364	358	772	14.79%	35.76%	41.35%
		SEMANTICALLY-RELATED	2421	185	761	548	740	478	30.57%	50.17%	60.93%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	192	855	625	608	550	25.11%	44.25%	56.75%
		SEMANTICALLY-RELATED	2421	184	740	543	766	463	31.64%	51.17%	61.83%
	Baseline		**2421**	**427**	**1994**	**0**	**0**	**1206**	**49.81%**	**60.48%**	**82.36%**

One by One Algorithm

Ww. size	MORPH. AWARENESS	LEX. RELTY.	W	f	d	p	C_{-d}	C_{+d}	R	P	C_v
3	SEMANTIC	NON- LEXICAL	2421	255	783	118	714	294	29.49%	51.63%	57.13%
	LEXICAL	SYNONYMOUS	2421	245	669	93	525	254	21.69%	34.84%	62.25%
		SEMANTICALLY-RELATED	2421	164	223	185	1010	53	41.72%	49.66%	84.01%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	184	292	174	872	285	36.02%	44.83%	80.34%
		SEMANTICALLY-RELATED	2421	159	207	181	1019	226	42.09%	49.59%	84.88%
								42			
5	SEMANTIC	NON- LEXICAL	2421	197	514	294	860	165	35.52%	50.29%	70.63%
	LEXICAL	SYNONYMOUS	2421	206	423	239	642	151	26.52%	35.83%	74.02%
		SEMANTICALLY-RELATED	2421	146	97	370	1097	23	45.31%	50.37%	89.96%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	148	133	371	947	231	39.12%	44.25%	88.39%
		SEMANTICALLY-RELATED	2421	142	83	365	1113	184	45.97%	50.68%	90.71%
								47			
7	SEMANTIC	NON- LEXICAL	2421	190	444	445	904	149	37.34%	50.59%	73.81%
	LEXICAL	SYNONYMOUS	2421	191	380	323	670	144	27.67%	36.22%	76.41%
		SEMANTICALLY-RELATED	2421	146	98	436	1092	19	45.11%	50.16%	89.92%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	151	122	475	940	240	38.83%	43.76%	88.72%
		SEMANTICALLY-RELATED	2421	144	88	431	1103	187	45.56%	50.39%	90.42%
								58			
11	SEMANTIC	NON- LEXICAL	2421	177	434	577	897	146	37.05%	49.56%	74.76%
	LEXICAL	SYNONYMOUS	2421	184	394	409	683	158	28.21%	37.06%	76.13%
		SEMANTICALLY-RELATED	2421	145	113	477	1085	23	44.82%	50.16%	89.34%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	149	119	566	950	116	39.24%	44.12%	88.93%
		SEMANTICALLY-RELATED	2421	141	105	474	1090	85	45.02%	50.11%	89.84%
	Baseline		**2421**	**427**	**1994**	**0**	**0**	**1206**	**49.81%**	**60.48%**	**82.36%**

One by One Algorithm with Fast Alternatives

Ww. size	MORPH. AWARENESS	LEX. RELTY.	W	f	d	p	C_{-d}	C_{+d}	R	P	C_v
3	SEMANTIC	NON- LEXICAL	2421	210	510	216	831	318	34.32%	48.85%	70.26%
	LEXICAL	SYNONYMOUS	2421	205	347	254	725	229	29.95%	38.79%	77.20%
		SEMANTICALLY-RELATED	2421	152	135	319	1015	81	41.92%	47.56%	88.15%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	167	181	322	917	107	37.88%	44.24%	85.63%
		SEMANTICALLY-RELATED	2421	152	136	339	1017	77	42.01%	47.68%	88.10%
5	SEMANTIC	NON- LEXICAL	2421	172	234	440	933	163	38.54%	46.30%	83.23%
	LEXICAL	SYNONYMOUS	2421	167	141	498	862	98	35.61%	40.80%	87.28%
		SEMANTICALLY-RELATED	2421	142	34	570	1073	22	44.32%	47.80%	92.73%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	144	47	552	989	30	40.85%	44.35%	92.11%
		SEMANTICALLY-RELATED	2421	142	31	564	1071	20	44.24%	47.64%	92.85%
7	SEMANTIC	NON- LEXICAL	2421	167	193	555	963	143	39.78%	46.72%	85.13%
	LEXICAL	SYNONYMOUS	2421	160	90	585	908	60	37.51%	41.82%	89.67%
		SEMANTICALLY-RELATED	2421	148	30	643	1082	20	44.69%	48.24%	92.65%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	149	38	662	988	24	40.81%	44.23%	92.28%
		SEMANTICALLY-RELATED	2421	147	28	634	1076	20	44.44%	47.91%	92.77%
			0	0	0	0	0	0			
11	SEMANTIC	NON- LEXICAL	2421	170	175	685	973	123	40.19%	46.87%	85.75%
	LEXICAL	SYNONYMOUS	2421	170	97	628	910	69	37.59%	42.25%	88.97%
		SEMANTICALLY-RELATED	2421	155	36	731	1052	29	43.45%	47.17%	92.11%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	162	40	741	988	30	40.81%	44.52%	91.66%
		SEMANTICALLY-RELATED	2421	151	34	734	1056	27	43.62%	47.23%	92.36%
17	SEMANTIC	NON- LEXICAL	2421	168	174	742	1007	122	41.59%	48.44%	85.87%
	LEXICAL	SYNONYMOUS	2421	177	83	668	898	61	37.09%	41.55%	89.26%
		SEMANTICALLY-RELATED	2421	164	37	796	1057	31	43.66%	47.61%	91.70%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	165	46	739	987	37	40.77%	44.66%	91.28%
		SEMANTICALLY-RELATED	2421	166	33	789	1061	27	43.82%	47.75%	91.78%
29	SEMANTIC	NON- LEXICAL	2421	197	177	761	967	127	39.94%	47.24%	84.55%
	LEXICAL	SYNONYMOUS	2421	202	116	704	872	82	36.02%	41.46%	86.86%
		SEMANTICALLY-RELATED	2421	193	65	797	1028	50	42.46%	47.53%	89.34%
	MORPHO-SEMANTIC	SYNONYMOUS	2421	197	62	770	948	42	39.16%	43.85%	89.30%
		SEMANTICALLY-RELATED	2421	189	63	808	1029	47	42.50%	47.44%	89.59%
	Baseline		**2421**	**427**	**1994**	**0**	**0**	**1206**	**49.81%**	**60.48%**	**82.36%**

Appendix 64

Mappings from *Claws* POS tags to the POSes of traditional grammar

Claws tag	POS	Notes on unmapped items (from BNC documentation available on licence from http://www.natcorp.ox.ac.uk/)
AJ0	ADJECTIVE	
AJC	ADJECTIVE	
AJS	ADJECTIVE	
AT0	ADJECTIVE	
AV0	ADVERB	
AVP	ADVERB	
AVQ	ADVERB	
CJC	CONJUNCTION	
CJS	CONJUNCTION	
CJT	CONJUNCTION	
CRD	ADJECTIVE	
DPS	ADJECTIVE	
DT0	PRONOUN	
DTQ	PRONOUN	
EX0	ADVERB	
ITJ	INTERJECTION	
NN0	NOUN	
NN1	NOUN	
NN2	NOUN	
NP0	NOUN	
ORD	ADJECTIVE	
PNI	PRONOUN	
PNP	PRONOUN	
PNQ	PRONOUN	
PNX	PRONOUN	
POS	NULL	The possessive or genitive marker *'s* or *'*
PRF	PREPOSITION	
PRP	PREPOSITION	
PUL	NULL	Punctuation mark
PUN	NULL	Punctuation mark
PUR	NULL	Punctuation mark
TO0	PREPOSITION	
UNC	NULL	Unclassified items which are not appropriately considered as items of the English lexicon.
VBB	VERB	
VBD	VERB	
VBG	VERB	
VBI	VERB	
VBN	VERB	
VBZ	VERB	
VDB	VERB	
VDD	VERB	
VDG	VERB	
VDI	VERB	

Claws tag	POS	Notes on unmapped items (from BNC documentation available on licence from http://www.natcorp.ox.ac.uk/)
VDN	VERB	
VDZ	VERB	
VHB	VERB	
VHD	VERB	
VHG	VERB	
VHI	VERB	
VHN	VERB	
VHZ	VERB	
VM0	VERB	
VVB	VERB	
VVD	VERB	
VVG	VERB	
VVI	VERB	
VVN	VERB	
VVZ	VERB	
XX0	ADVERB	
ZZ0	NULL	Alphabetical symbols (e.g. *A*, *a*, *B*, *b*, *c*, *d*)

Appendix 65

The WordNet model

Further details of some individual classes can be found in Appendix 1.

The WordNet model was implemented in Java using the NetBeans 6.0.1 Integrated Development Environment, from www.netbeans.org. This IDE was used to monitor the behaviour of the classes developed and scenarios which provoked exceptions and to implement further functionality throughout the project. The data sources were the WordNet Prolog files downloaded from http://wordnet.princeton.edu/obtain. Synsets, word senses and relations are represented in the model as instances of corresponding Java classes (Class Diagrams 1 and 2 represent the original version of the model). The model is constructed from the Prolog files, by the constructor of the `NaturalLanguageProcessor`, which in turn invokes the `Wordnet` constructor, which instantiates the synsets. The object-oriented design was intended to facilitate extensions and deletions, rendering the model suitable for correction and enrichment of WordNet.

Synset instantiation (*Class Diagrams 1, 2 & 3*)

An empty global synset map is created[25].

A subclass of `WordSense` is created from each record in file *wn_s.pl*. This record includes a synset type field corresponding to one of the 5 subclasses of `Synset`: `NounSynset`, `VerbSynset`, `AdjectiveClusterHead`, `AdjectiveSatellite` or `AdverbSynset`. The `WordSense` created will be a `Noun`, `Verb`, `Adjective` or `Adverb`

[25] `Map<Integer, Synset>`

as implied by the synset type field. If an entry exists in the global synset map for the synset ID specified in the record, then this `Synset` is retrieved from the global synset map, otherwise the specified subclass of `Synset` is created, and is added to the global synset map, indexed by the synset ID. The `WordSense` created is inserted into the `List<WordSense>` encapsulated in the `Synset` at the position specified by the word number field in the record[26].

The WordNet sense keys are read from file *wn_sk.pl*. Each record in this file specifies a Synset ID, a word number and a sense key. The corresponding `Synset` is retrieved from the global synset map and the corresponding `WordSense` is retrieved from the `List<WordSense>` encapsulated in the `Synset`. The sense key is broken down into its components, as specified by the WordNet documentation and these are stored in separate fields of the `WordSense`.

The WordNet glosses are read from file *wn_g.pl*. These are broken down into their logical components which may include one or more glosses, one or more examples and one or more attributions of those examples. These are stored in separate fields of the corresponding `Synset`, the attributions being co-indexed to the corresponding examples. This was achieved by reverse engineering the format in which the glosses are stored in the Prolog records.

Encoding the WordNet Relations (*Class Diagrams 4 & 5*)

With the exception of file *wn_fr.pl*, all the remaining files in the download specify WordNet relations which hold between synsets or between word senses, or occasionally between a synset and a word sense. The names of these files specify the `Relation.Type` of the `WordnetRelation` The records in the files comprise 2, 4 or 5 fields. In all cases 2 fields specify the source and target synsets between which the relation holds. Where the relation holds between word senses, 2 further fields specify the source and target word numbers. In the case of CLASS_MEMBER relations, a fifth field specifies the subtype of the relation. Zero as a word number for either source or target indicates that the source or target of a relation which normally holds between word senses is exceptionally a whole synset. Any other word number specifies an individual word sense. Some relations can only hold between certain subclasses of `Synset` and `WordSense`.[27]

In the model, relations are held within their source objects in a relations map.[28] These maps are created when the objects are instantiated, at which point their set of possible relation types is fixed. Every time a Relation is encoded, it is added to the `Set<Relation>` mapped to by its `Relation.Type` and its converse is added to the `Set<Relation>` mapped to by the converse type (Appendix 22) in the target object. Identifiers for both source and target are encapsulated in every `Relation`. The target of every `WordnetRelation` is represented as the corresponding Synset ID, and the

[26] As there are no zero-valued word numbers in the Prolog files, the word number is decremented by 1, so that word number 1 is at index 0 in the `List`.
[27] This information is held in static fields of the corresponding classes.
[28] `Map<WordnetBuilder.Relation.Type, Set<Relation>>` inherited by classes `Synset` and `WordSense` from abstract class `WordWrapper`.

target word of every `WordSenseRelation` (`WordnetRelation` holding between word senses) is held as the corresponding word number.

Adding Sentence frames

If specified by a Boolean parameter passed to the `NaturalLanguageProcessor` constructor, the 35 `WordNetVerbFrame` objects are instantiated and stored in a `MutableCollection`. The assignations of frames to verbs are read from file *wn_fr.pl*. Each record in this file holds a synset ID, a word number and a frame number. Zero as a word number indicates that the frame number is to be assigned to an entire `VerbSynset`; any other word number specifies an individual `Verb` within that `VerbSynset`. To facilitate the interrogation of the frame information, they are all assigned to an individual `Verb`. Where a `VerbSynset` is specified, the frame is assigned to every `Verb` within that `VerbSynset`.

Building the Lexicon (*Class Diagrams 2 & 7*)

In the original model the main dictionary was implemented as a `Map<String, LexicalRecord>` where each `LexicalRecord`, corresponding to a single word form, held a sense map[29] mapping from the synset ID of every Synset containing the corresponding word form to the relevant `LexicalInformationTuple`, holding the sense number, the word number and the tag count of a single `WordSense`.

In the original implementation, The `Lexicon` constructor created an empty main dictionary and iterated through the global synset map and through the word sense list of every `Synset`. It looked up the word form of every `WordSense` in the main dictionary and retrieved the corresponding `LexicalRecord`, or created a new one with the corresponding mapping if no entry was found. In either case a new entry was added to the sense map, mapping from the ID of the current `Synset` to a new `LexicalInformationTuple`, whose word number is determined from the current index in the word sense list and whose other fields are obtained from the `WordSense`.

The `Lexicon` constructor was subsequently modified to match the modified design (§§1.3.2.4, 3.5.3) which accommodates POS-specific queries. The modified constructor retrieves the `GeneralLexicalRecord` corresponding to the `WordSense`, or creates a new one. The sense map of a `GeneralLexicalRecord` is a `Map<Wordnet.PartOfSpeech, POSSpecificLexicalRecord>` from which the `POSSpecificLexicalRecord` corresponding to the POS of the current `Synset` must be retrieved. If there is no corresponding entry in the sense map of the `GeneralLexicalRecord`, then a new `POSSpecificLexicalRecord` must be created along with the required mapping. The sense map of a `POSSpecificLexicalRecord` is as described in the previous paragraph.

Initialising the Lemmatiser (*Class Diagram 6*)

The lemmatiser requires two maps, one for regular inflections and one for exceptions (Class Diagram 6). In the regular inflection map[30], each lemmatisable word ending for

[29] `Map<Integer, LexicalInformationTuple>`

[30] `Map<Wordnet.PartOfSpeech, Map<String, POSTaggedMorpheme[]>>`

each POS maps to an array of one or more possible lemmas. The lemmas are POS-tagged because mappings are required from lemmatisable word endings to lemmas belonging to a different POS, mainly because there are numerous adverbs in "-ly" which are not encoded as word senses in WordNet. This map was originally based on the table to be found in the WordNet documentation at http://wordnet.princeton.edu/man/morphy.7WN.[31] This data proved to be incomplete and has been extended as and when items missing from the table came to light[32]. The regular inflection map has been constructed in such a way that the correct mapping will always be the first encountered (for instance the mapping "ches" to "ch" is encountered before the mapping "es" to "e".

Each entry in the exception map[33] maps from a whole word, with its POS specified, to an `IrregularStemPair` which encapsulates a POS and a maximum of 2 irregular stems. It is populated from the four WordNet exception files available with the download (*noun.exc; verb.exc; adj.exc; adv.exc*), to which a few items have been added.[34]

The Lemmatiser services lemmatisation queries, by first looking up the whole word in the regular inflection map and then searching for the longest lemmatisable ending which corresponds to the end of the word for which there is an entry in the regular inflection map. A single most probable lemma or a number of possible lemmas may be returned depending on how the query is specified. An array of inflectional suffixes (§1.3.2.5) which occur preceded by an apostrophe may also be consulted[35].

[31] As the size of the data was very small it was hard-coded into the Lemmatiser constructor.
[32] but the constructor has not, as yet, been modified to read this data from a file.
[33] `Map<POSTaggedWord, IrregularStemPair>`
[34] hard-coded
[35] One or more hard-coded verbs will be returned.