Making Sense of Roget

Jeremy Ellman[1] and Robert Bell[1]

Abstract

Being able to resolve word senses could improve precision in Information Retrieval, machine translation or other natural language applications. However, automatic disambiguation rarely provides benefits in proportion to its costs.

In this study we have tried to determine the dominant or 'most frequent' sense of an ambiguous word. By identifying this, future applications could statistically choose a 'correct' word sense the majority of the time without computation.

We investigated several ambiguous nouns and the frequency of their co-occurring synonyms from Roget's thesaurus using the British National Corpus. This indicates the likelihood of the possible word senses. The results have been evaluated with some success against varying lexical resources including WordNet and the Concise Oxford Dictionary.

1 Introduction

Ambiguity is defined as the possibility of interpreting an expression in more than one way. Thus, saying that word is ambiguous indicates that it has more than one sense, or meaning.

All natural languages, including English, consist of words that have a varying number of senses or definitions. For example, the Collins dictionary states that the word 'line' has fifty senses whilst the word 'damage' has three.

When an ambiguous word is used in conversation or written in a book the human listener or reader naturally understands its intended sense. The same cannot be said of computers.

Computers are not able to distinguish between the different senses of the word. When considering Internet searching, if a query has a large degree of ambiguity then it may not be possible to find for what one is searching.

The idea of automatically disambiguating words for electronic applications has been of interest to researchers since the 1950s (Sanderson 2000, Stokoe et al. 2003). The objective of Word Sense Disambiguation (WSD) is well known as an important issue in Natural Language Processing (NLP) applications.

There are numerous applications of WSD, such as Information Retrieval and Machine Translation. Information Retrieval (IR) is the study of searching, indexing and retrieving documents or text from a collection. Up until the early 1990's it was a limited research domain with interest mainly to librarians and information experts (Baeza-Yates and Ribeiro-Neto 1999). However the origin of the World Wide Web

[1] School of Computing, Engineering, and Information Sciences, Northumbria University
e-mail: Jeremy.Ellman @northumbria.ac.ukz

spawned a greatly enlarged research field and information retrieval now covers a vast search space.

Machine Translation (MT) applications translate text from one language to another while eliminating ambiguity in the translation. Preiss and Stevenson (2004) explain this by discussing an English-French MT system that needs to establish whether the word 'bank' is defined as 'financial institution' or the 'verge of a river' and must elect to translate this as 'banque' or 'bord'. MT is consequently a active application area of Word Sense Disambiguation.

Recent work on Word Sense Disambiguation in English has focused on WordNet (Fellbaum 1998) as the sense inventory. However, Navigli (2006) notes that WordNet uses sense distinctions that are too fine, making WSD harder than need be.

An alternative to WordNet is Roget's thesaurus. Roget is a well-known source of inspiration and an aid to writers in search of a synonym or a particular expression to express subtle nuances of meaning. It is possibly less well known that Roget is arranged into a balanced hierarchical structure that is typically six levels deep with approximately one thousand categories that are subdivided by part of speech.

Roget's thesaurus is then a potentially useful tool for language analysis. Indeed, Roget was suggest by Morris and Hirst (1991) as suitable for the construction of lexical chains and was used as such by Ellman (2000, 2007)

Word sense ambiguity is a problem when using Roget as an analytic tool. That is, approximately half of the words in Roget appear in multiple categories, as they are ambiguous. Furthermore, there is no indication as to which of the senses are more or less common. Thus for example, Roget gives no clue that 'plane' is more common as a mode of transport, than as a smooth surface. Whilst word sense disambiguation algorithms can help, these often perform better if a preferred, or most common, sense is known. Typically, dictionaries such as the OED, Collins, etc will list their entries according to either the lexicographers' perceptions as to priority, or, increasingly according to usage in a large text corpus.

Consequently, one obstacle to the wider use of Roget in language applications is the inability to identify preferred senses of words, which is the subject of this paper.

1.1 Experimental Hypothesis

Firth (1957) stated that a word can be known from the company it keeps. This is no truer than with Roget's thesaurus since that does not encode any relationship more specific than association. That is, each Roget category contains a collection of words that may be synonyms, antonyms or related terms, and the resource itself gives few clues as to which.

We may hypothesise however that by considering a sufficient number of loose associations in a large corpus, such as the British National Corpus (BNC), we will find that members of a thesaurus category often occur together in a text for stylistic and descriptive reasons. This co-occurrence frequency could indicate the intended meaning of a polysemous word. Furthermore, the more preferred a polysemous sense, the more frequent will be its associations with other members of that thesaural category.

The problem with this hypothesis is that each word whose preferred sense we wish to determine will share its categories with other ambiguous words. Thus, 'bank' shares its entries with 'interest'. There are two possible resolutions to this dilemma. Firstly, we could exclude all polysemous words from consideration that share the

Roget category with our target word. This however brings with it a sparse data problem: There are too few words remaining to give accurate co-occurrence statistics. Our resolution to this problem is to introduce the notion of a 'polysemous disambiguator'.

A polysemous disambiguator is word that shares a Roget category with a polysemous word, but neither of the alternate senses of both words share Roget categories. For example, 'worm' is associated with 'slug' as a small mollusc, but the other senses of worm (e.g. 'bad-person', 'disease-infection') are not associated with any other senses of slug ('drinking-draught', 'ammunition').

1.2 Outline of the Paper

This paper proceeds as follows: Firstly we will describe the experimental methodology. This includes the derivation of a gold standard set of human judgements for a common set of polysemous terms that will be used to test the hypothesis. Next we describe the algorithm used to derive co-occurrence frequency information from the BNC. The experimental results are then analysed and followed with conclusions and future work.

2 Method

The main objective of this investigation is to automatically determine preferred senses of ambiguous words using Roget's thesaurus as the sense inventory. We hypothesise that this is possible if we exploit the frequency distribution of co-occurring words from the same Roget categories using the BNC as a source of frequency information.

A manually created gold standard benchmark will be created to evaluate the results. If the most frequent sense of a word from the BNC matches the most frequent sense from the gold standard then the hypothesis will be supported and it may be possible to determine the dominant sense of a word using this unsupervised technique. Furthermore, as a corollary, if the preferred sense can be identified then it may also be possible to identify an ordering of sense from most to least preferred.

There are several steps to achieving this. These are:

- The sample set of ambiguous words
- Creating a gold standard benchmark
- Obtaining the co-occurring terms from Roget
- Searching the BNC
- The Sense Preference Algorithm

2.1 The sample set of ambiguous words

To test our hypothesis, we need a set of polysemous words. Differing degrees of polysemy are important here. For example, a word such as 'hide' that has two senses in WordNet would be easier to disambiguate than a word such as 'bank' that has ten.

Although the BNC is large text resource it is still subject to the sparse data problem. That is, although some word pairs are plausible neighbours according to their identical category membership in Roget, they do not co-occur in the BNC.

Consequently, we selected used Yarowsky's (1992) sample set of twelve polysemous words.

Star	Taste
Mole	Interest
Galley	Issue
Cone	Duty
Bass	Sentence
Bow	Slug

Table 1: The test set of polysemous word

Now that a word sample set has been selected, we need to identify their senses and create a gold standard benchmark. That is, a list of definitions based on multiple lexical resources.

2.2 Creating a gold standard Benchmark

The gold standard is an ordered list of senses intended as benchmark to evaluate the senses preference algorithm. It was created prior to the experiment to avoid accidental bias by gathering various lexical resources and manually combining their word sense definitions. This gives a definitive, ordered, sense list that that is independent of any particular resource or inventory.

The language resources used were Collins Concise dictionary (Collins 2001), Concise Oxford Dictionary (Pearsall 2001) and WordNet (Fellbaum 1998). Like Roget, WordNet is organised in a hierarchical structure. However, it is an American resource that may have dialectic variations on the sense definitions offered by the two British English dictionaries. It should also be noted that WordNet was not compiled by professional lexicographers so this could also influence the resource comparison (Seo 2004).

To demonstrate how a benchmark is created first the definitions need to be taken from the resources and summarised into short descriptions and listed together so that comparisons can be observed.

As an example, consider the word 'galley'. Table 2 below shows the definitions from which the gold standard sense order will be hand picked.

SENSE	COLLINS	WORDNET	OXFORD
1	An oared or sailed vessel	An oared or sailed vessel	An oared or sailed vessel
2	A ship or aircraft's kitchen	A sea going vessel	A ship or aircraft's kitchen
3	Long Rowing boat	An aircraft's kitchen	A Printing type tray
4	A Printing type tray	A ship's kitchen	

Table 2: 'Galley' sense definitions

From comparing the three resources a benchmark can be derived for this word. Some discretion must be used when comparing resources as some use definitions not found in others. Also note that WordNet has a tendency to repeat previous senses or give similar definitions to prior senses. In this case WordNet sense 2 is similar to sense 1, and also Collins sense 3 is similar to sense 1. For this reason senses can be combined manually to give a more general definition. For example, the gold standard word ‘galley’ is shown below (Table 3)

SENSE	DEFINITION
1	An oared or sailed vessel
2	A ship or aircraft’s kitchen
3	A Printing type tray

Table 3: ‘Galley’ gold standard

2.3 Roget categories and co-occurring synonyms

Roget’s thesaurus is one of the most distinguished lexical resources. It differs from other thesauri as it uses a classification system and is not just a “synonym dictionary”. As mentioned, Roget’s classification system uses a hierarchical structure of classes, sections, heads, paragraphs and keywords. The higher in the structure concepts are, the more general the categories. The highest upper level categories consist of abstractions such as Abstract, Volition and Space, whereas the lower in the structure the more specific the topics become.

Looking at the gold standard, we shall first consider the word ‘duty’ (see Table 4).

DUTY		
Sense No.	Definition	Roget Category No.
1	Moral or legal obligation	4728, 1952, 2134
2	Public payment tax	1156
3	Measure of engine effectiveness	3625

Table 4: Duty gold standard with Roget categories

The task now is to list the sense definitions for the word ‘Duty’ and hand pick the categories that correspond to the gold standard. The sense categories for ‘duty’ are shown in Table 5 below.

Table 5 shows the complete set of noun sense categories in Roget for *‘Duty’*. The Roget categories corresponding to the gold standard definitions need to be hand selected from these. In gathering ‘a moral or legal obligation’ (sense 1 in benchmark) of ‘duty’ the co-occurring synonyms occur in Roget categories 4728, 1952, and 2134. Each one of these categories holds a list of synonym words with respect to the first sense definition of the gold standard. These synonyms will be tested with the sample word ‘duty’ within the BNC for frequencies.

One problem this sample demonstrates is errors of omission: That is, when a gold standard definition is not included in the Roget classifications. This happens with the third sense from the gold standard, ‘measure of engine effectiveness’. There is no a Roget Category that relates to this sense. Therefore, that sense must be excluded from further consideration.

3161-Abstract_relations-Causation-Liability-liability-undef-noun
2837-Abstract_relations-Quantity-Bond-connecting_medium-bond-noun
1294-Emotion-religion_and_morality-Interpersonal_emotion-Courtesy-courteous_act-noun
597-Emotion-religion_and_morality-Morality-Dueness-dueness-noun
4728-Emotion-religion_and_morality-Morality-Duty-duty-noun
541-Emotion-religion_and_morality-Morality-Respect-respects-noun
1952-Emotion-religion_and_morality-Morality-Right-right-noun
1172-Emotion-religion_and_morality-Religion-Ritual-church_service-noun
209-Emotion-religion_and_morality-Religion-Worship-cult-noun
2134-Volition-The_exercise_of_the_will-Conditional_social_volition-Promise-promise-noun
3756-Volition-The_exercise_of_the_will-General_social_volition-Obedience-obedience-noun
1156-Volition-The_exercise_of_the_will-Possessive_relations-Price-tax-noun
40-Volition-The_exercise_of_the_will-Prospective_volition-Business-function-noun
41-Volition-The_exercise_of_the_will-Prospective_volition-Business-job-noun
100-Volition-The_exercise_of_the_will-Prospective_volition-Requirement-needfulness-noun
3626-Volition-The_exercise_of_the_will-Volition_in_general-Motive-motive-noun
812-Volition-The_exercise_of_the_will-Volition_in_general-Necessity-necessity-noun
1914-Volition-The_exercise_of_the_will-Voluntary_action-Exertion-labour-noun

Table 5: Duty sense categories in Roget thesaurus

2.4 Searching the BNC

The word senses and their co-occurring category members once derived from Roget can be queried in the BNC to collect co-occurrence frequency statistics. The BNC uses its own query language known as the Corpus Query Language (CQL). CQL queries can include the following useful parameters:

- Word window size
- Part of Speech data
- Specifying accents and special characters

The word window size lets you select words within a set limit. An example may be to find how many times ‘cat’ is followed by ‘dog’ within ten words of each other (ten being the size of the word window). If the word window was increased to

fifty there may be more occurrences, with possibly a greater number of spurious associations. That query would be represented:

- `cat*dog/50`

Part of Speech enables using queries that only return a particular sense of a word such as a noun or a verb. If the following was inserted as a query:

- `can=NN1`

This would only return the singular common noun for the word 'can'.

Users can also search using specific accents and special characters. An example would be in searching for the place name 'Zürich'. In order to achieve this, the query is enclosed with double quotes to clarify that a special character is attached in the query.

For this work the only style of query that needs to be generated for this study is shown below.

```
(duty=NN1)#(excise)/50
```

Where # represents an AND operator and the number fifty represents the size of the word window within which the two words can co-occur.

The word window size of fifty was based on that of Preiss (2004). Preiss (2004) also references earlier work from Gale, Church and Yarowsky (1992) who adopted a word window size of fifty. We determined experimentally that any limit lower than fifty encountered the sparse data problem discussed previously, whereas any larger window size reduced precision due to spurious associations.

2.5 The Sense Preference Algorithm

Sense preference is calculated using word co-occurrence frequency information from the BNC. Here we are looking for a target word occurring together with other members of its categories in Roget's thesaurus. This needs to take account of differing word frequencies, and the variable number of words in each category.

The disparity in frequencies is corrected for using the Dice Coefficient, which is a standard measure of association from Information Retrieval (e.g. Evert 2004, Baeza-Yates and Ribeiro-Neto 1999). Thus, the association between words w_i, and w_j with a BNC frequency function f is given by:

$$Dice(w_i, w_j) = \frac{2 \times f(w_i, w_j)}{f(w_i) + f(w_j)}$$

Equation 1: The Dice Coefficient

Difference in length of Roget categories is accounted for by taking the arithmetic mean of the Dice coefficients. Since some word pairs do not appear in the BNC (e.g. 'bow' co-occurs with 'violin' but not 'catgut') these are excluded from N,

the size of the Roget category. Thus the Sense_Frequency of a word w in category c is given by equation 2:

$$Sense\_Frequency(w_c) \approx \frac{\sum_{w_j \in C}(Dice(w_c, w_j))}{N}$$

$$where\left\langle (w_c \in C) \,\&\, (w_j \in C) \,\&\, (w_i \neq w_j) \,\&\, \neg\exists(C')[(w_c \in C') \,\&\, (w_j \in C')]\right\rangle$$
$$\&\, N = \left|C\right| \,\&\, Dice(w_c, w_j) \neq 0$$

Equation 2: Sense Frequency

Equation 2 also includes the proviso that w_j should either be monosemous, or, if polysemous, should not share another category with w_c.

The senses preference order of any ambiguous word W is given by the following Sense Preference function.

Function Senses_Preferences(Word W)

i. **`Identify`** all Roget categories R for Word W
ii. $Senses_W = 0$;
iii. For Each Category C in R
iii.1.1 **`Insert`** W, C, **`Sense_Frequency`**(W_c) into $Senses_W$
iv. End For
v. Return **`Sort`**($Senses_W$))
vi. End Function

At this point we have sample word of ambiguous words, and have shown how the frequency of any pairs of words may be found from the BNC.

3 Results

The principal objective of this study is to determine where a preferred sense can be identified by reference to frequency distributions from the BNC. This can be simply accomplished by examining which sense is in the preferred, first, position.

We shall also test the corollary that the algorithm should indicate an ordering of the less frequent senses. This may be tested by putting the senses into order, and comparing the ranking with that of the gold standard. Here we will use the Spearman rank order statistic, which is nonparametric and so makes no assumptions about how the variable values are distributed.

3.1 General results of Sample word set

Table 6 below shows the sense rankings for the gold standard and thse determined algorithmically for the sample word set. To clarify, consider ‘taste’. The first sense in the gold standard is indeed the ranked first or most frequent in the BNC output (proving a dominant sense), however the second sense in the gold standard is only ranked 4th most frequent in the BNC output. Also shown is the Spearman’s rho which

indicates the correlation between the rankings (an asterisk indicates a statistically significant correlation).

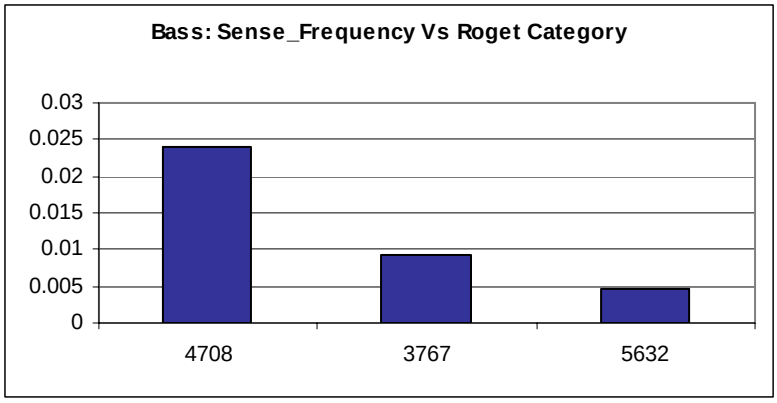
Sample Word	gold standard	Algorithm Output	Spearman's ρ		gold standard	Algorithm Output	Spearman's ρ
Bass	1	1	0.5	Mole	1	3	0.521
	2	3			2	1,4	
	3	2			3	2	
Sentence	1	1	0.447		4		
	2,	2,4,5			5	5	
	3	3		Issue	1		-0.1
Slug	1	1	0.849*		2	4	
	2	2			3	1	
	3				4	5	
	4	4			5	2	
	5	3			6	3	
Cone	1	1	1.0*	Bow	1	4	-0.078
	2	2			2	2	
					3	3	
Taste	1	1,2	0.458		4	5, 7	
	2	5			5	6	
	3	6			6		
	4	3			7	1	
	5	4		Duty	1	1,3,4	-02582
Star	1	2	0.798		2	2	
	2	3			3		
	3			Galley	1	4	-0.316
	4	1			2	1,2	
	5				3	3	
	6	6,4		Interest	1	5	-0.9
	7	5			2	3	
	8	7			3	4	
					4	2	
					5	1	
					6		

Table 6: Sample word set frequency comparison

3.2 Results for Individual Words

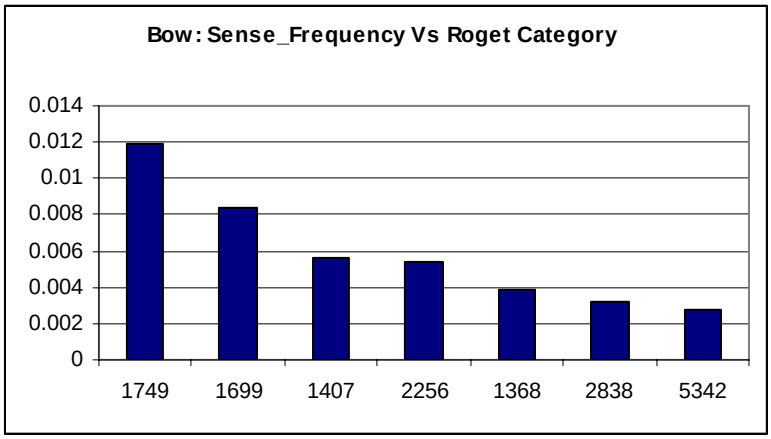
In this section we show the sense ordering for the sample words individually. This is of interest since we can see the identity of the preferred senses, and analyse the rank ordering in table 6 more precisely. That is, minor differences between two values alter the rank correlation just as much as large differences. However, a minor difference means that less weight should be given to the correlation statistics.

In the following graphs 1-11, Appendix A shows the equivalence between the Roget category numeric values to the gold standard meanings.



Roget sense 4708 is the lowest male singing voice. The fish sense, 5632, is shown as least preferred.

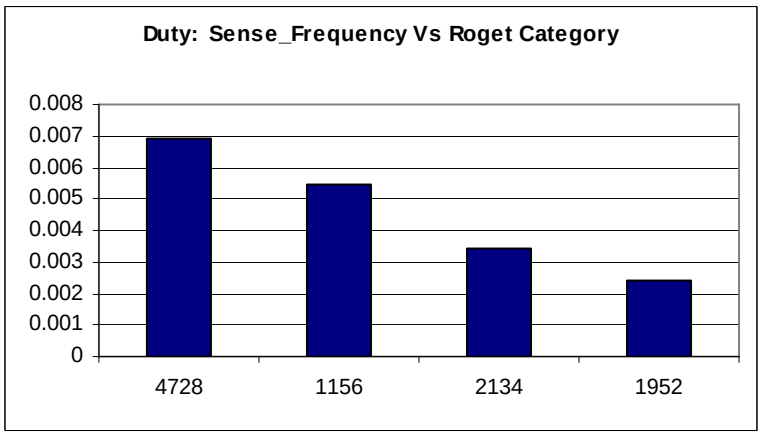
Graph 1: Bass Sense Frequency



The primary sense here is the font end of a boat, followed by the arrow device.

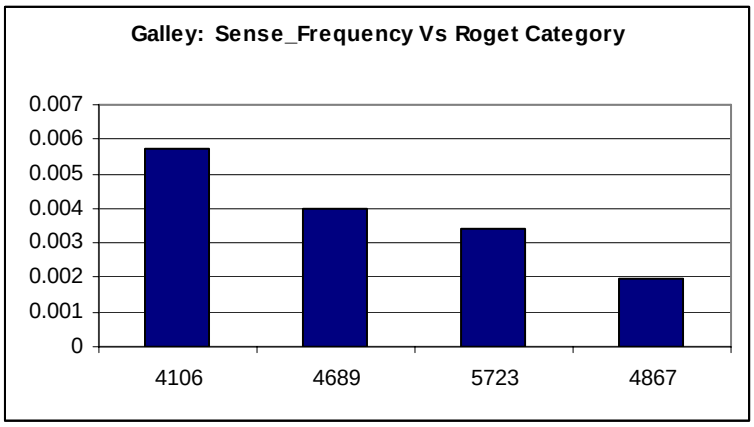
Note the relatively minor differences in sense frequency for the remaining senses, curve, greeting, and slip knot.

Graph 2: Bow Sense Frequency



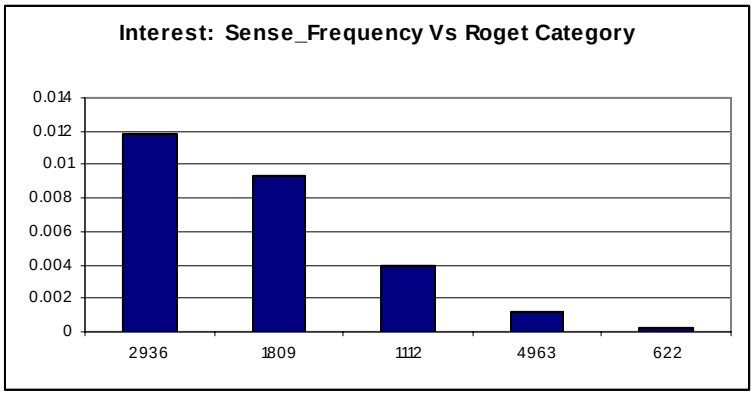
The primary senses for duty is moral obligation, followed by tax.

Graph 3: Duty Sense Frequency



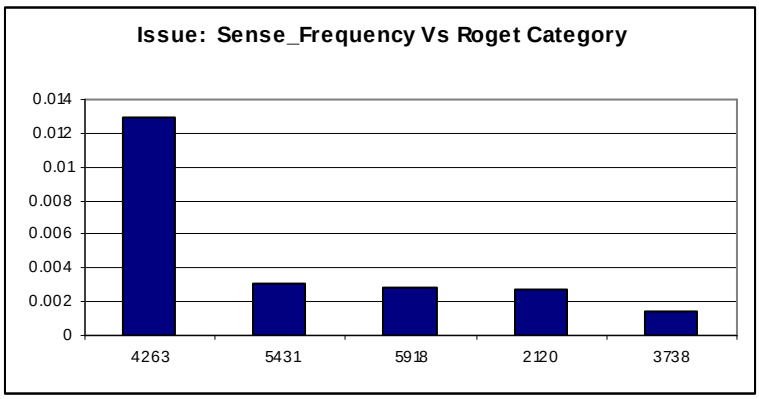
Graph 4: Galley Sense Frequency

The ship, or aircraft kitchen, sense dominates here. Intuitively this is far more appealing for a modern corpus such as the BNC, although the dictionaries consider the sailing vessel meaning to be preferred.



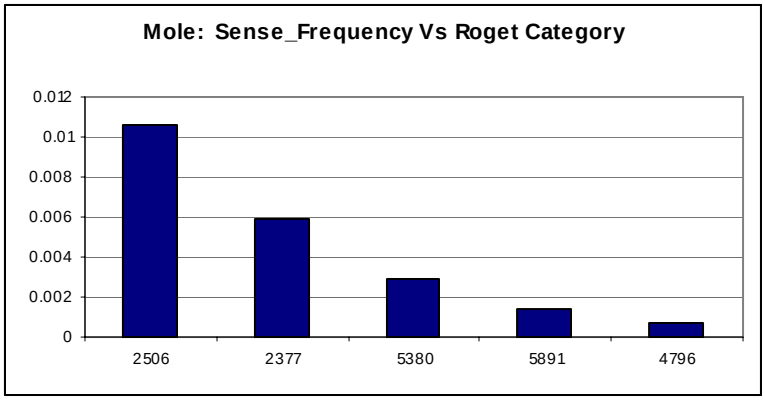
Graph 5: Interest Sense Frequency

The algorithm indicates that the financial sense of interest is preferred, whilst that of curiosity is quite rare.



Graph 6: Interest Sense Frequency

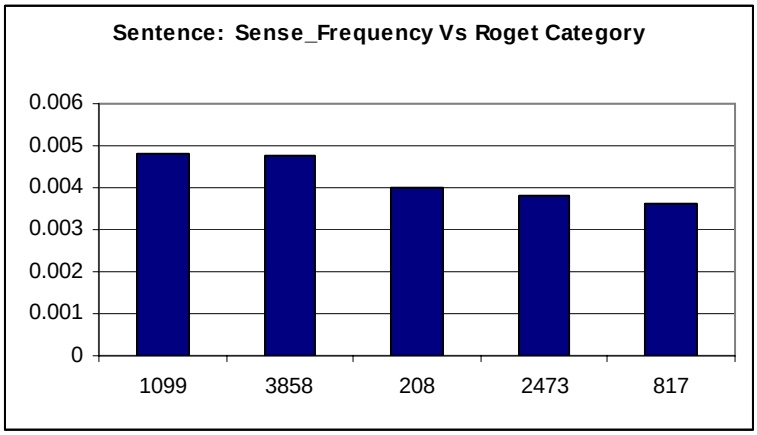
The preferred sense of issue is ‘a point in question’. There is little difference between the remaining senses and the lack of correlation between the ranks should therefore be considered unreliable.



Graph 7: Mole Sense Frequency

The automatically determined primary sense of mole is a skin blemish, followed by the burrowing animal.

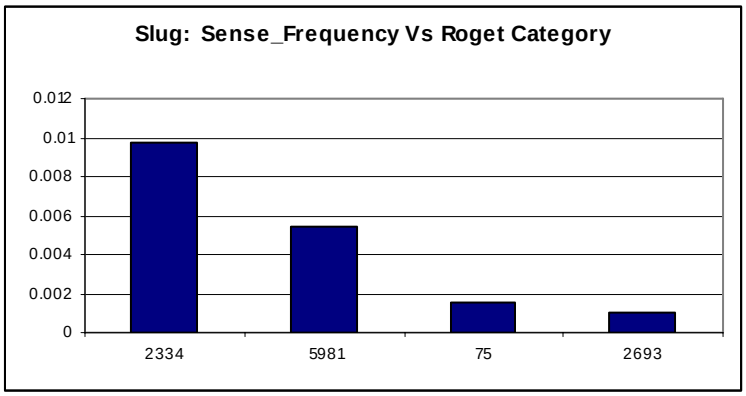
This is the same ordering as in the Oxford Illustrated Dictionary, but inverted in the Concise Oxford.



Graph 8: Sentence Sense Frequency

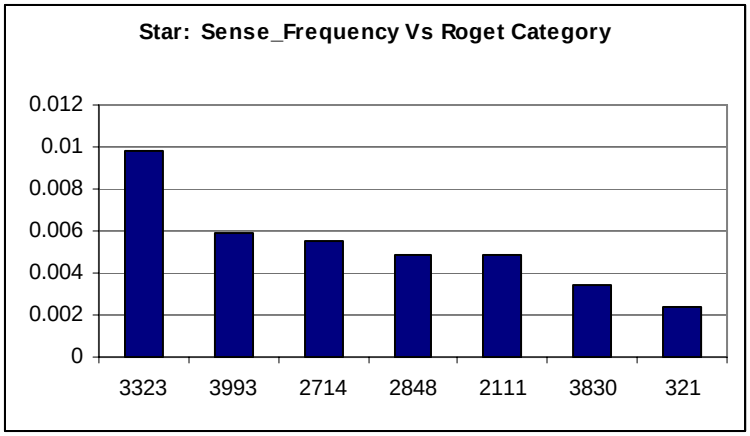
The primary sense of sentence is that of a set of words, followed closely by the legal decision.

The difference here between the values is small, and consequent ranking is questionable.



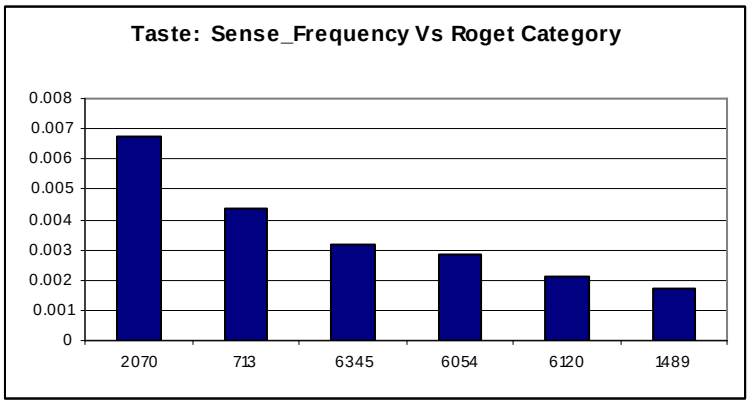
Graph 9: Slug Sense Frequency

The algorithm identifies slug as primarily a small mollusc, followed by the bullet sense.



Graph 10: Star Sense Frequency

Star is preferred here as a pointed figure. The following two senses (hot gaseous mass and celestial body) would dominate if combined. Otherwise, there is again little difference in preference for the following senses.



Graph 11: Taste Sense Frequency

The first two senses are amalgamated in the gold standard as the sensation of taste. Preference and aesthetic discernment follow.

3.3 Evaluation and Discussion

From Table 6 it can be seen that the algorithm correctly identified the preferred sense in the gold standard for 50% of the sample word set. That is, for *Bass*, *Cone*, *Duty*, *Sentence*, *Slug* and *Taste*. The probability of this being a chance result is less than 1/1350. Where the algorithm identified a different result to the gold standard (such as with issue, mole, and interest) there is some suspicion that the gold standard did not reflect current use (see section 3.2 above).

Furthermore, for the sample as a whole there were eighteen out of fifty seven (32%) senses where there was a match between output sense rank and that defined in the gold standard. The words *Bass* and *Sentence* had a complete correlation between BNC output and gold standard for each word sense. There was a wide range of coefficient ranging from the perfect positive correlation of *Bass* and *Sentence*, but strong negative correlation for the word *Interest*.

Looking at the results there are several points that should be considered. Firstly, the gold standard benchmark relied heavily on human judgement. Thus, some sense definitions were similar and the decision was taken to merge these. An example of this was the word 'Cone' where originally the first two senses read the following:

Sense 1 - Solid figure with circular plane base

Sense 2 – Similar cone shape tapered to a point

In determining definition rankings there were occasions when one resource in the gold standard (Collins, WordNet or Concise Oxford Dictionary) would have a

sense definition with a high rank that other resources did not have on there sense lists. The decision had to be taken as to whether one high-ranking definition deserves to be in the gold standard when it not mentioned in the other two resources.

WordNet has the tendency to repeat or generate very similar senses that need to be combined. WordNet also brings American usages. In the 'mole' definitions for example, the third sense for WordNet is 'spicy sauce'. Neither the Concise Oxford Dictionary nor Collins contain this definition so the dilemma is to consider whether this definition is included instead of the 'breakwater' definition that has lower ranks but is in all three resources.

Although broad, Roget does not include all possible categories. For example, 'slug' has a third ranking definition as 'a unit of mass for acceleration[2]' in the gold standard. However Roget does not have an entry in this sense. In this case, this sense needs to be excluded from the rank correlation.

Finally, it should be noted that Roget often includes more entries (i.e. senses) of a word than contained in the gold standard. For example 'Duty' (table 5) has eighteen entries in Roget, whilst the standard (Appendix A) contains three. Even allowing for several Roget categories being matched to one standard entry there are still deficiencies with the standard.

4 Related Work

This work is most closely related to that in Word Sense Disambiguation (WSD), which has been ably summarized by Ide and Véronis (1998), Preiss and Stevenson (2004) and many others. WSD is an active research area that has been the subject of many recent workshops, most notably three Senseval workshops (Kilgarriff and Palmer 2000, Kilgarriff and Rosenzweig (2000), Edmonds and Kilgarriff 2002, Mihalcea and Edmonds 2004) and the recent re-incarnation as Semeval (Agirre et al. 2007).

More specifically, McCarthy et al. (2004), have described a technique to identify predominant word senses in untagged text. Their technique used an automatically generated thesaurus, and a similarity measurement derived from WordNet (Pedersen et al. 2004). That measure is based SemCor, the subset of the Brown corpus that has been manually tagged with WordNet senses. Consequently, McCarthy et al. (2004) will inherit inaccuracies due to the manual tagging in SemCor, and other, known deficiencies in the WordNet sense hierarchy.

Finally, Yarowsky (1992) reported a word sense disambiguation system based on Roget's thesaurus. His system produced statistical models of the Roget categories, and used these to probabilistically disambiguate the set of ambiguous words in text. Yarowsky (1992) differs from this work as here we are looking to identify preferred senses of words rather than disambiguate.

[2] 'Obsolete unit of mass, equal to 14.6 kg/32.17 lb. It is the mass that will have an acceleration of one foot per second when under a force of one pound weight.' http://www.thefreedictionary.com/

5 Conclusion and Further Work

This paper has described a completely unsupervised method for determining default senses for Roget's thesaurus using word frequency data from the British National Corpus. The method consists of taking terms that unambiguously share a thesaural category with a target word, and calculating the Dice Coefficient as a measure of association. Once normalised for category size, this give a sense preference ordering. Most importantly, the first member of the ordered senses indicates a default, or preferred sense that will be useful in NLP, IR, and MT applications.

The method has been evaluated against a hand created gold standard sense inventory of twelve polysemous words for which preferred senses were correctly recognised for six (50%). This gold standard encountered problems common to any attempt to matching sense inventories. That is, senses exist in one resource, but not the other, categories overlap, and have differing divergences. Problems with the gold standard notwithstanding, the evaluation was positive when considering ranking of sense preferences, and the primary aim of identifying a preferred sense.

When considering further work, it is a simple extension to determine preferred senses for the complete Roget. The question would be how this could be evaluated. We propose the best evaluation here would be to apply the output to an extrinsic task. That is, a problem for which human judgements exists as a baseline, and where a program's performance would be assessed based on improved knowledge of preferred senses. An example application here would be Web person search (e.g. See Ellman and Emery 2007) as that application uses large data volumes and thesaural based similarity matching.

Although Roget's thesaurus has been the focus of this paper, the technique described is generic and would also be applicable to other hierarchical taxonomies or classification systems such as the European Common Procurement Vocabulary (CPV), or the International Patent Classification (IPC) (Fall et al. 2003).

Acknowledgement

We are most grateful to Addison Wesley Longman Limited for permission to use extracts of *The Original Roget's Thesaurus of English Words and Phrases* Copyright © 1987 by Longman Group UK Ltd. in portions of this work.

References

Agirre Eneko, Lluís Màrquez, and Richard Wicentowski 2007 Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007) Prague, June 23–24th 2007

Baeza-Yates R and Ribeiro-Neto B (1999) Modern Information Retrieval, Addison-Wesley

Collins Concise Dictionary (2001) Collins; 5Rev Ed edition ISBN-10: 0007109784

Ellman Jeremy and Gary Emery 2007 UNN-WePS: Web Person Search using co-Present Names and Lexical Chains. Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007) Association for Computational Linguistics Prague June 2007

Evert Stefan 2004 Association measures Available at http://www.collocations.de/AM/ [accessed July 2007]

Fall, C. J., Törcsvári, A., Benzineb, K., and Karetka, G. 2003. Automated categorization in the international patent classification. *SIGIR Forum* 37, 1 (Apr. 2003)

Fellbaum, Christiane. 1998 WordNet - An electronic lexical database, MIT Press, Cambridge, Massachusetts and London, England, 1998

Firth, J. (1957). Papers in linguistics. London: Oxford University Press

Ide Nancy, Jean Véronis, 1998 Introduction to the special issue on word sense disambiguation: the state of the art, Computational Linguistics, v.24 n.1, March 1998

Kilgarriff Adam and Palmer Martha (2000) Introduction to the Special Issue on Senseval, Computers and the Humanities Vol. 34: pp. 1–13

Kilgarriff Adam and Rosenzweig Joshua (2000) Framework and results for English Senseval, Computers and the Humanities Vol. 34: pp. 15–48

McCarthy, D., Koeling, R., Weeds, J., and Carroll, J. 2004. Finding predominant word senses in untagged text. In *Proceedings of the 42nd Annual Meeting on Association For Computational Linguistics* (Barcelona, Spain, July 21–26, 2004). Annual Meeting of the ACL. Association for Computational Linguistics,

Mihalcea Rada, and Phil Edmonds (2004) 'Proceedings of Senseval-3: The Third International Workshop on the Evaluation of Systems for the Semantic Analysis of Text' Barcelona, Spain, July 25–26, 2004 available online at http://www.senseval.org/senseval3/proceedings [accessed 04 July 2007]

Morris, Jane, and Graeme Hirst, (1991). Lexical cohesion computed by thesaural relations as an indicator of the structure of text. Computational Linguistics, 17(1), 21–48.

Navigli, Roberto 2006. Meaningful Clustering of Senses Helps Boost Word Sense Disambiguation Performance, Proc. of the 44th Annual Meeting of the Association for Computational Linguistics joint with the 21st International Conference on Computational Linguistics (COLING-ACL 2006), Sydney, Australia, July 17–21st, 2006, pp. 105–112.

Pearsall Judy (ed.) 2001 The Concise Oxford Dictionary Press; 10Rev Ed edition (12 Jul 2001) ISBN-10: 019860436X

Pedersen Ted, Siddharth Patwardhan, Jason Michelizzi: 2004 WordNet: Similarity - Measuring the Relatedness of Concepts. proc. AAAI 2004: 1024–25.

Philip Edmonds and Adam Kilgarriff 2002 Introduction to the special issue on evaluating word sense disambiguation systems. Natural Language Engineering 8 (4): pp. 279–91.

Preiss Judita (2004) Probabilistic Word Sense Disambiguation. Computer speech and Language 18 (2004) 319–37

Preiss Judita and Mark Stevenson (2004) Introduction to the special issue on word sense disambiguation. Computer Speech and Language 18 pp. 201–207.

Sanderson Mark and C J Van Rijsbergen (2000) The Impact on Retrieval Effectiveness of Skewed Frequency Distributions. ACM Transactions on Information Systems Vol.17 No.4 pp. 440–65

Sanderson Mark (2000) Retrieving with Good Sense Information Retrieval Vol.2(1), pp. 49–69

Seo HC et al. (2004) Unsupervised word sense disambiguation using WordNet relatives. Computer Speech and Language 18 pp. 253–73

Stokoe Chris et al. (2003) Word Sense Disambiguation in Information Retrieval Revisited, proc ACM SIGIR July 2003 P: 159–166

Yarowsky David (1992) Word sense disambiguation using statistical models of Roget's categories trained on large corpora, In proceedings of COLING Conference pp. 454–60

Appendix A: The gold standard listings

Note: Blank entries indicate that Roget does not contain an entry for the term in that sense.

BASS		
Sense No.	**Definition**	**Roget Category No.**
1	Lowest male singing voice	4708
2	Lowest musical range	5632
3	A fish	3767

BOW		
Sense No.	**Definition**	**Roget Category No.**
1	A slip-knot	2256
2	Arrow device	1699
3	Violin bow	1407
4	Curve or bend	5342, 1368
5	Head greeting	2838
6	Retreat (Bow out)	
7	Fore end of a boat	1749

CONE		
Sense No.	**Definition**	**Roget Category No.**
1	A solid cone shaped figure	3774
2	Dry fruit of a conifer	3965
3	Retinal cone of the eye	

DUTY		
Sense No.	**Definition**	**Roget Category No.**
1	Moral or legal obligation	4728, 1952, 2134
2	Public payment tax	1156
3	Measure of engine effectiveness	

GALLEY		
Sense No.	**Definition**	**Roget Category No.**
1	Oared or sailed vessel	416, 4867
2	Ship or aircraft kitchen	4689, 4106
3	Printing tray	5723

INTEREST		
Sense No.	**Definition**	**Roget Category No.**
1	Curiosity	622
2	A Hobby	1112
3	Advantage	4963
4	Financial Money lent	1809
5	Share or financial stake in	2936
6	A party or group concern	

ISSUE		
Sense No.	**Definition**	**Roget Category No.**
1	Distribution	
2	Periodical	2120
3	A point in question	4263
4	Outlet	3738
5	Law children, progeny	5431
6	Outcome	5918

MOLE		
Sense No.	**Definition**	**Roget Category No.**
1	Infiltrator	5380
2	Skin blemish	2506, 5891
3	burrowing mammal	2377
4	Molecular weight	
5	Breakwater protection	4796

SENTENCE		
Sense No.	**Definition**	**Roget Category No.**
1	Set of words	1099
2	Law decision	817, 2473, 3858
3	Logic series	208

SLUG		
Sense No.	**Definition**	**Roget Category No.**
1	Small mollusc	2334
2	Bullet	5981
3	Unit of mass	
4	Printing spacing	2693
5	A tot of liquor	75

STAR		
Sense No.	**Definition**	**Roget Category No.**
1	Celestial body	3993
2	Hot gaseous mass	2714
3	Astrology	
4	Emblem figure with 5 points	3323
5	Blaze on an animal	
6	A glamorous celebrity	2848, 3830
7	An asterisk	321
8	A headliner	2111

TASTE		
Sense No.	**Definition**	**Roget Category No.**
1	Taste sensation	713, 2070
2	Small sample of food/drink	6120
3	A slight experience	1489
4	Preference	6345
5	Aesthetic discernment	6054