Automatic Thesaurus Construction

Dongqiang Yang | David M. Powers
School of Informatics and Engineering
Flinders University of South Australia
PO Box 2100, Adelaide 5001, South Australia
Dongqiang.Yang|David.Powers@flinders.edu.au

Abstract[1]

In this paper we introduce a novel method of automating thesauri using syntactically constrained distributional similarity. With respect to syntactically conditioned co-occurrences, most popular approaches to automatic thesaurus construction simply ignore the salience of grammatical relations and effectively merge them into one united 'context'. We distinguish semantic differences of each syntactic dependency and propose to generate thesauri through word overlapping across major types of grammatical relations. The encouraging results show that our proposal can build automatic thesauri with significantly higher precision than the traditional methods.

Keywords: syntactic dependency, distribution, similarity.

1 Introduction

The usual way of automatic thesaurus construction is to extract the top *n* words in the similar word list of each seed word as its thesaurus entries, after calculating and ranking distributional similarity between the seed word and all of the other words occurring in the corpora. The attractive aspect of automatically constructing or extending lexical resources rests clearly on its time efficiency and effectiveness in contrast to the time-consuming and outdated publication of manually compiled lexicons. Its application mainly includes constructing domain-oriented thesauri for automatic keyword indexing and document classification in Information Retrieval, Question Answering, Word Sense Disambiguation, and Word Sense Induction.

As the ground of automatic thesaurus construction, distributional similarity is often calculated in the high-dimensional vector space model (VSM). With respect to the *basic elements* in VSM (Lowe, 2001), the dimensionality of word space can be syntactically conditioned (i.e. grammatical relations) or unconditioned (i.e. '*a bag of words*'). Under these two context settings, different similarity methods have been widely surveyed, for example for '*a bag of words*' (Sahlgren, 2006) and for grammatical relations (Curran, 2003; Weeds, 2003). Moreover, the framework conducted by Padó and Lapata (2007) compared the difference between the two settings. They observed that the syntactically constrained VSM outperformed the unconditioned one that exclusively counts word co-occurrences in a ±*n* window.

Given the hypothesis that similar words share similar grammatical relationships and semantic contents, the basic procedure for estimating such distributional similarity can consist of (1) pre-processing sentences in the corpora with shallow or complete parsing; (2) extracting syntactic dependencies into distinctive subsets or vector spaces (**X**s) according to head-modifier, including adjective-noun (**AN**) and adverb or the nominal head in a prepositional phrase to verb (**RV**) and grammatical roles including subject-verb (**SV**) and verb-object (**VO**); and (3) determining distributional similarity using similarity measures such as the Jaccard coefficient and the *cosine*, or probabilistic measures such as KL divergence and information radius. On the other hand, without the premise of grammatical relations in sematic regulation, calculating distributional similarity can simply work on word co-occurrences.

Instead of arguing the pros and cons of these two context representations in specific applications, we focus on how to effectively and efficiently produce automatic thesauri with syntactically conditioned co-occurrences.

Without distinguishing the latent differences of grammatical relations in dominating word meanings in context, most approaches simply chained or clumped these syntactic dependencies into one unified context representation for computing distributional similarity such as in automatic thesaurus construction (Hirschman et al., 1975; Hindle, 1990; Grefenstette, 1992; Lin, 1998; Curran, 2003), along with in Word Sense Disambiguation (Yarowsky, 1993; Lin, 1997; Resnik, 1997), word sense induction (Pantel and Lin, 2002), and finding the predominant sense (McCarthy et al., 2004). These approaches improved the distributional representation of a word through a fine-grained context that can filter out the unrelated or unnecessary words produced in the traditional way of '*a bag of words*' or the unordered context, given that the parsing errors introduced are acceptable or negligible.

It is clear that these approaches, based on observed events, often scaled each grammatical relation through its frequency statistics in computing distributional similarity, for example in the weighted (Grefenstette, 1992) or mutual information based (Lin, 1998) Jaccard coefficient.

Although they proposed to replace the unordered context with the syntactically conditioned one, they have overlooked the linguistic specificity of grammatical relations in word distribution. Except for the extraction of syntactically conditioned contexts, they in fact make no differentiation between them, which are similar to computing distributional similarity with unordered context. The advantage of using the syntactic constrained context has not yet been fully exploited when yielding statistical semantics from word distributions.

To fully harvest the advantages of computing distributional similarity in the syntactically constrained contexts, we proposed to first categorize contexts in terms of grammatical relations, and then overlapped the top n similar words yielded in each context to generate automatic thesauri. This is in contrast to averaging distributional similarity across these contexts, which is commonly adopted in the literature.

2 Context interchangeability of similar words

Word meaning can be regarded as a function of word distribution within different contexts in the form of co-occurrent frequencies, where similar words share similar contexts (Harris, 1985). Miller and Charles (1991) propose that word similarity depends on to what extent they are interchangeable across different context settings. The flexibility of one word or phrase substituting another indicates its extent to be synonymous providing that the alternation of meaning in discourse is acceptable. We calculated distributional similarity in different syntactic dependencies such as subject-predicate and predicate-object. Given the interchangeability of synonyms or near-synonyms in different contexts, semantically similar words derived with distributional similarity should span at least two types of syntactically constrained contexts. In other words, once we can derive the thesaurus items from each dependency set, the final thesaurus comprises the intersection of the items across any two types of dependency sets.

The heuristic of deriving automatic thesauri with the interchangeability of synonyms or near-synonyms in contexts (‘***any two***’) can be expressed:

- Nouns: $\bigcup_{i,j}(S_i \cap S_j)$ where i and j stand for any two types of dependency sets in terms of grammatical relations: **AN**, **SV**, and **VO**.
- Verbs: $\bigcup_{i,j}(S_i \cap S_j)$ where i and j stand for any two of **RV**, **SV**, and **VO**.

where for a given word, S is the thesaurus items produced through distributional similarity in a single dependency set. Note that we also used the heuristics of ‘***any three***’ and ‘***any four***’ to construct automatic thesauri, but found most target words had no distributionally similar words under these stricter conditions than ‘***any two***’. We did not attempt to demonstrate the conditions here.

We similarly hypothesized the union of all grammatical relations from the co-occurrence matrices as a baseline (‘***all***’), which compute distributional similarity with the union of all relations and can be indicated:

- Nouns: $S_{\bigcup_i}$ where i is one of **AN**, **SV**, and **VO**
- Verbs: $S_{\bigcup_i}$ where i is one of **RV**, **SV**, and **VO**

3 Syntactically constrained distributional similarity

To automate thesauri, we first employed an English syntactic parser based on Link Grammar to construct a syntactically constrained VSM. The word space consists of four major syntactic dependency sets that are widely adopted in the current research on distributional similarity. Following the reduction of dimensionality on the dependency sets, we created the latent semantic representation of words through which distributional similarity can be measured so that thesaurus items can be retrieved.

3.1 Syntactic dependency

The syntactically conditioned representation mainly rely on the following grounds: (1) the meaning of a noun depends on its modifiers such as adjectives, nouns, and the nominal head in a prepositional phrase as well as the grammatical role of a noun in a sentence as a subject or object (Hirschman et al., 1975; Hindle, 1990); and (2) the meaning of a verb depends on its direct object, subject, or modifier such as the head of a prepositional phrase (Hirschman et al., 1975). These results are partly consistent with the findings in studying word association and the psychological reality of the paradigmatic relationships of WordNet (Fellbaum, 1998).

With the hypothesis of ‘one sense per collocation’ in WSD, Yarowsky (1993) observed that the direct object of a verb played a more dominant role than its subject, whereas a noun acquired more credits for disambiguation from its nominal or adjective modifiers. As an application of the distributional features of words, Resnik (1997) and Lin (1997) employed the selectional restraints in subject-verb, verb-object, head-modifier and the like to conduct sense disambiguation.

The syntactic dependencies can provide a clue for tracking down the meaning of a word in context. Cruse (1986) points out that the semantic requirements are of two directions in head-modifier and head-complement, namely, determination (selector and selectee) and dependency (dependee and depender). The determination requirement emphasizes the dominant role of the selector in the semantic traits of a construction, while the dependency supplements some additional traits to formulate the integrity of the construction.

3.2 Categorizing syntactic dependencies

Suppose that a tuple $<w_i, r, w_j>$ describes the words: w_i and w_j, and their bi-directional dependency relation r. For example, if w_i modifies w_j through r, all such w_j with r to w_i form a context profile for w_i, likewise w_i for w_j. In the hierarchy of syntactic dependencies (Carroll et al., 1998), the major types of grammatical relationships (r) can be generally clustered into:

- **RV**: verbs with all verb-modifying adverbs and the head nouns in the prepositional phrases;
- **AN**: nouns with noun-modifiers including adjective use and pre/post-modification;
- **SV**: grammatical subjects and their predicates;
- **VO**: predicates and their objects.

To capture these dependencies we employ a widely used and freely available parser[2] based on Link Grammar (Sleator and Temperley, 1991). In Link Grammar each word is equipped with ‘left-pointing’ and/or ‘right-pointing’ connectors. Based on the crafted rules of the connectors in validating word usages, a link between two words can be formed in reflecting a dependency relation. Apart from these word rules, ‘crossing-links’ and ‘connectivity’ are the two global rules working on interlinks, which respectively restrict a link from starting or ending in the middle of pre-existed links and force all the words of a sentence to be traced along links. There are in total 107 major link types in the Link Grammar parser (ver. 4.1), whereas there are also various sub-link types that specify special cases of dependencies. Using this parser, we extracted and classified the following link types into the four main types of dependencies:

- **RV**

1. *E*: verbs and their adverb pre-modifiers
2. *EE*: adverbs and their adverb pre-modifiers
3. *MV*: verbs and their post-modifiers such as adverbs, prepositional phrase

- **AN**

1. *A*: nouns and their adjective pre-modifiers
2. *AN*: nouns and their noun pre-modifiers
3. *GN*: proper nouns and their common nouns
4. *M*: nouns and their various post-modifiers such as prepositional phrases, adjectives, and participles

- **SV**

1. *S*: subject-nouns/gerunds and their finite verbs. There are also some sub-link types under S, for example, Ss*g stands for gerunds and their predicates, and Sp plural nouns and their plural verbs
2. *SI*: the inversion of subjects and their verbs in questions

- **VO**

1. *O*: verbs and their direct or indirect objects
2. *OD*: verbs and their distance-complement
3. *OT*: verbs and their time objects
4. *P*: verbs and their complements such as adjectives and passive participles

Note that except for **RV**, we define the **AN**, **SV**, and **VO** dependencies almost identically to shallow parsers (Grefenstette, 1992; Curran, 2003), or a full parser of MINIPAR (Lin, 1998) but we retrieve them instead through the Link Grammar parser.

Consider, for example, a short sentence from British National Corpus (BNC):

> *‘Home care Coordinator, Margaret Gillies, currently has a team of 20 volunteers from a variety of churches providing practical help to a number of clients already referred.’*

The parse of this sentence with the lowest cost in the link grammar parser is shown in Figure 1, where LEFT-WALL indicates the start of the sentence

```
    +----------------------------------------------------------------------
    |                        +-------------------Ss------------------+
    |                        +---------MX---------+                  +-----
    +-----------Wd-----------+      +------Xd-----+                  +---Os
    |        +--AN--+---GN---+      |     +---G---+-Xc-+     +---E---+   +-
    |        |      |        |      |     |       |    |     |       |   |
LEFT-WALL Home.n care.n Coordinator , Margaret Gillies , currently has.v a

-------------------------------Xp------------------------------------
----------MVp---------------+                                 +------
---+    +----Jp----+        +---Jp---+                        +------
Ds-+-Mp-+  +--Dmcn-+        |  +-Dsu-+--Mp--+--Jp--+----Mg----+
|    |  |       |        |  |     |      |      |          |
team.n of 20 volunteers.n from a variety.n of churches.n providing.v

-----------------------------------------------------------------+
                                                                 |
------MVp-----------+                                            |
----Os---------+    +---J---+           +--------Mv--------+     |
     +----A----+    | +--Ds-+--Mp-+--Jp-+        +----E----+     |
     |         |    | |     |     |     |        |         |     |
practical.a help.n to a number.n of clients.n already referred.v .
```

Figure 1: A complete linkage of parsing a sentence using Link Grammar

The parse of this sentence with the lowest cost in the link grammar parser is shown in Figure 1, where LEFT-WALL indicates the start of the sentence. We can classify four types of grammatical relations from this parse, namely:

- **RV**: <*currently, E, has*>, <*already, E, referred*>
- **AN**: <*home, AN, care*>, <*care, GN, coordinator*>, <*volunteer, Mp, team*>, <*church, Mp, variety*>, <*practical, A, help*>, <*client, Mp, number*>, <*referred, Mv, clients*>
- **SV**: <*coordinator, Ss, has*>
- **VO**: <*has, Os, team*>, <*providing, Os, help*>

After parsing the 100 million-word BNC and filtering out non-content words and morphology analysis, we separately extracted the relationships to construct four parallel matrixes or co-occurrence sets, denoted as R_X: $\mathbf{RV}_X$, $\mathbf{AN}_X$, $\mathbf{SV}_X$, and $\mathbf{VO}_X$ in terms of the four types of syntactic dependencies above. The row vectors of R_X denoted respectively $\mathbf{Rv}_X$, $\mathbf{An}_X$, $\mathbf{Sv}_X$, and $\mathbf{Vo}_X$ for the four dependencies. Similarly, the column vectors of R_X are denoted as $\mathbf{rV}_X$, $\mathbf{aN}_X$, $\mathbf{sV}_X$, and $\mathbf{vO}_X$ respectively.

Consider $\mathbf{SV}_X$ a m by n matrix representing subject-verb dependencies between m subjects and n verbs. We illustrate the **SV** relation using the rows ($\mathbf{Sv}_X$ or $\{X_{i,*}\}$) of $\mathbf{SV}_X$ corresponding to nouns conditioned as subjects of verbs in sentences, and the columns ($\mathbf{sV}_X$ or $\{X_{*,j}\}$) to

[2] http://www.link.cs.cmu.edu/link/

verbs conditioned by nouns as subjects. The cell $X_{i,j}$ shows the frequency of the *i*th subject with the *j*th verb. The *i*th row $X_{i,*}$ of $\mathbf{SV}_X$ is a profile of the *i*th subject in terms of its all verbs and the *j*th column $X_{*,j}$ of $\mathbf{SV}_X$ profiles the *j*th verb versus its subjects.

The parsing results are shown in Table 1, where *Dim* refer to the size of each matrix in the form of rows by columns, and *Freq* segmentations are the classification of frequency distribution, and Token/Type stands for the statistical frequencies of specific relationships with their corresponding dependency category R.

Dim \ *Freq*			1	2-10	11-20	21-30	>31
$\mathbf{AN}_X$	48.5 by	Token	1,813.7	6,243.4	1,483.1	799.8	3,617.8
	37.6	Type	1,813.7	2,040.0	103.6	32.2	44.9
$\mathbf{RV}_X$	37.4 by	Token	863.1	2,276.4	481.4	234.9	692.2
	14.2	Type	863.1	751.9	33.8	9.5	10.9
$\mathbf{SV}_X$	32.7 by	Token	511.8	1,699.4	297.8	133.3	380.7
	11.3	Type	511.8	587.4	21.0	5.4	6.0
$\mathbf{VO}_X$	6.1 by	Token	488.5	1,811.5	475.4	266.2	1,286.9
	33.3	Type	488.5	575.1	33.1	10.7	15.6

Table 1: The statistics of the syntactically conditioned matrices derived from parsing BNC (thousand)

Given different methodologies to implementing parsing, it is hardly fair to appraise a syntactic parser. Molla and Hutchinson (2003) compared the Link Grammar parser and the Conexor Functional Dependency Grammar (CFDG) parser with respect to intrinsic and extrinsic evaluations. In the intrinsic evaluation the performance of the two parsers was compared and measured in terms of the precision and recall of extracting four types of dependencies, including subject-verb, verb-object, head-modifier, and head-complement. In the extrinsic evaluation a question-answering application was used to contrast the two parsers. Although the Link Grammar parser is inferior to the CFDG parser in locating the four types of dependencies, they are not significantly different when applied in question answering. Given that our main task is to investigate the function of the syntactic dependencies: **RV**, **AN**, **SV**, and **VO**, acquired with the same Link Grammar parser, in automatic thesaurus construction, it is appropriate to use the Link Grammar parser to extract these dependencies.

3.3 Dimensionality reduction in VSM

The four syntactically conditioned matrices, as shown in Table 1, are extremely sparse with nulls in over 95% of the cells. Instead of eliminating the cells with lower frequencies, we kept all co-occurrences unchanged to avoid worsening data sparseness.

Our matrices record the context with both syntactic dependencies and semantic content. These dual constraints yield rarer events than word co-occurrences in '*a bag of words*'. However, they impose more accurate or meaningful grammatical relationships between words providing the parser is reasonable accurate.

We initially substituted each cell frequency $freq(X_{i,j})$ with its information form using $log(freq(X_{i,j})+1)$ to retain sparsity ($0 \rightarrow 0$) (Landauer and Dumais, 1997). It can produce 'a kind of space effect' that can lessen the gradient of the frequency-rank curve in Zipf's Law (1965), reducing the gap between rarer events and frequent ones.

Singular Value Decomposition (SVD) often acts as an effective way of reducing the dimensionality of word space in natural language processing. A reduced SVD representation can diminish both 'noise' and redundancy whilst retaining the useful information that has the maximum variance. This approach has been dubbed Latent Semantic Analysis (LSA) (Deerwester et al., 1990; Landauer and Dumais, 1997) and maps the word-by-document space into word-by-concept and document-by-concept spaces. Note that the 'noisy' data in the raw co-occurrence matrices mainly comes from the results of wrong parsing and also redundancy exists as a common problem of expressing similar concepts in synonyms.

Typically at least 200 principal components are employed in Information Retrieval to describe the SVD compressed word space. Instead of optimising the semantic space versus other algorithms (through tuning the number of principal components in applications or evaluations), we specified a fixed dimension size for the compressed semantic space, which is thus not expected to be optimal for our experiment. We established 250 as a fixed size of the compressed semantic space. Among the singular values, the first 20 components account for around 50% of the variance, and the first 250 components for over 75%.

As is usual with the SVD/LSA application, we assume that the semantic representation of words is a linear combination of eigenvectors representing their distinct subcategorizations and senses, and that relating the uncorrelated eigenvector feature sets of different words can thus score their proximity in the semantic space.

3.4 Distributional similarity

We consistently employed the cosine similarity of word vectors as used in LSA and commonly adopted in assessing distributional similarity (Salton and McGill, 1986; Schütze, 1992). The cosine of the angle θ between vectors x and y in the *n*-dimensional space is defined as:

$$\cos\theta = \frac{x \cdot y}{\|x\|\|y\|} = \frac{\sum_{i=1}^{n} x_i y_i}{\sqrt{\sum_{i=1}^{n} x_i^2 \sum_{i=1}^{n} y_i^2}}$$

where the length of x and y is ||x|| and ||y||.

Note that the accuracy and coverage of automatic term clustering inevitably depend on the size and domains of the corpora employed, as well as similarity measures. Consistently using one similarity method—the *cosine*, our main task in this paper is to explore the context interchangeability in automatic thesaurus construction, rather than to compare different similarity measures with one united syntactic structure that combines all the dependencies together. Although taking into account more similarity measures in the evaluations may solidify conclusions, this would take us beyond the scope of the work.

4 Evaluation

4.1 The 'gold standard' thesaurus

It is not a trivial task to evaluate automatic thesauri in the absence of a benchmark set. Subjective assessment on distributionally similar words seems a plausible approach to assessing the quality of term clusters. It is practically unfeasible to implement it given the size of the term clusters. A low agreement on word relatedness also exists between human subjects.

The alternative way of measuring term clusters is to contrast them with existing lexical resources. For example, Grefenstette (1993) evaluated his automatic thesaurus with a 'gold standard' dataset consisting of Roget's Thesaurus ver. 1911, Macquarie Thesaurus, and Webster's 7th dictionary. If two words were located under the same topic in Roget or Macquarie, or shared two or more terms in their definitions in the dictionary, they were counted as a successful hit for synonyms or semantic-relatedness. To improve the coverage of the 'gold standard' dataset, Curran (2003) incorporated more thesauri: Roget's Thesaurus (supplementing the free version of 1911 provided by Project Gutenberg with the modern version of Roget's Thesaurus II), Moby Thesaurus, The New Oxford Thesaurus of English, and The Macquarie Encyclopaedic Thesaurus.

The 'gold standard' datasets are not without problem due to their domain and coverage, because they are at best a snapshot of general or specific English vocabulary knowledge (Kilgarriff, 1997; Kilgarriff and Yallop, 2000). Moreover, the organization of thesauri forces different notions of being synonymous or similar, given the etymologic trend of words and different purposes of lexicographers. For example, as 1 of 1,000 topics in Roget's Thesaurus ver. 1911, there are two groups of synonyms {*teacher*, *trainer*, *instructor*, *institutor*, *master*, *tutor*, *director*, etc.} or {*professor*, *lecturer*, *reader*, etc.} under the topic of *teacher*. They express an academic concept of being in the position of supervision over somebody. In the noun taxonomy of WordNet, the synonym of *teacher* only consists of *instructor*, affiliated with the coordinate terms (sharing one common superordinate) such as *lecturer* and *reader*, or the hyponyms such as *coach* and *tutor*, or the hypernyms such as *educator* and *pedagogue*. As for *professor* and *master*, they both distance *teacher* by three links through their hypernym *educator*.

Subject to the availability of these thesauri or dictionaries, we incorporated both WordNet and Roget's Thesaurus, freely acquired, into the 'gold standard' thesaurus. WordNet only consists of paradigmatic relations and organizes a fine-grained semantic taxonomy mainly with the relationships of syn/antonym, IS-A, HAS-A, whereas Roget's Thesaurus covers both syntagmatic and paradigmatic relations and hierarchically clusters related words or phrases into each topic without explicitly annotating their relationships.

Kilgarriff and Yallop (2000) claimed that WordNet, along with the automatic thesauri generated under the hypothesis of similar words sharing similar syntactic structures, are *tighter* rather than *looser* in defining whether they are 'synonyms' or related words. This contrasts with Roget and the automatic thesauri derived through unordered word co-occurrences. Since we accounted for distributional similarity in the syntactically conditioned VSM, the reasonable way of evaluating it is to compare our automatic thesauri to WordNet. Apart from that, to perform a systematic evaluation on the relationships among distributionally similar words, we also included Roget as a supplement to the 'gold standard', as it covers words with both paradigmatic and syntagmatic relationships.

4.2 Similarity comparison

We defined two distinctive measures to compare automatic thesauri with the 'gold standard', which are Sim_{WN} for WordNet and Sim_{RT} for Roget.

4.2.1 Similarity in WordNet

Sim_{WN} is based on the taxonomic similarity method proposed by Yang and Powers (2005; 2006). Since Yang and Powers's method outperformed most popular similarity methods in terms of correlation with human similarity judgements, we employed them in the evaluation. Given two nominal or verbal concepts: $c1$ and $c2$, Sim_{WN} scores their similarity with:

$$Sim(c1, c2) = \alpha_{str} \times \alpha_t \times \beta_t^{dist-1}, dist \le \gamma$$

- α_{str}: 1 for nouns but for verbs successively falls back to α_{stm} the verb stem polysemy ignoring sense and form; or α_{der} the cognate noun hierarchy of the verb; or α_{gls} the definition of the verb.
- α_t: the path type factor to specify the weights of different link types, i.e. syn/antonym, hyper/ hyponym and holo/meronym in WordNet.
- β: the probability associated with a direct link between concepts (type t).
- *dist*: the distance between two concept nodes
- γ: the path length *dist* is limited to depth factor γ, otherwise the similarity is 0

As for multiple senses of a word, word similarity maximizes their sense or concept similarity in WordNet.

Yang and Powers (2005) compared their taxonomic similarity metric with human judgements on the 65 noun pairs, where the cut-off point 2.36 of human similarity scores for nouns on a Likert scale from 0 to 4 divides each dataset into similar ($\geq$ 2.36) and dissimilar subsets (< 2.36). We found that the cut-off of 2.36 for nouns corresponds to the searching depth limit γ = 4 in Sim_{WN}, and likewise the cut-off of 2 on the 130 verb pairs (Yang and Powers, 2006) corresponds to $\gamma = 2$. Thus for the noun candidates in automatic thesauri, we set up $\gamma = 4$, to identify similar words within the distance of less than four links. If two nodes are syn/antonyms or related to each other in the taxonomy with the shortest path length of less than 4, we counted them as a successful hit. So too is the shorter distance limit $\gamma = 2$ for verb candidates.

4.2.2 Similarity in Roget's Thesaurus

Roget's Thesaurus divides its hierarchy into seven levels from the top *class* to the bottom *topic*, and stores topic-related words under 1 of 1,000 topics. Sim_{RT} counted it a hit if two words are situated under the same *topic*.

Note that the relationships among the 'gold standard' words retrieved by Sim_{RT} are anonymous. Although WordNet only organizes paradigmatic relationships, Sim_{WN} does not distinguish in what way two words are similar, for example, IS-A, HAS-A, or a mixture of them, and only collects words within a distance from zero (syn/antonyms) to four links in WordNet.

4.3 Candidate words in the 'gold standard'

		WordNet						Roget	Total
		SA	D1	D2	D3	D4	∑		
Noun	**aN**$_X$	462	2,825	14,244	41,483	48,625	107,639	141,102	232,181
	An$_X$	458	2,887	14,278	41,940	49,267	108,830	142,218	234,424
	vO$_X$	439	2,619	13,027	37,433	43,620	97,138	133,733	214,727
	Sv$_X$	434	2,607	12,938	37,355	43,274	96,608	131,527	212,156
∑X		469	2,979	14,967	44,185	52,054	114,779	146,435	244,245
Verb	**rV**$_X$	1,282	24,702	58,617			84,601	81,713	144,545
	Vo$_X$	1,260	24,265	57,225			82,750	79,771	141,039
	sV$_X$	1,269	24,354	57,642			83,265	80,681	142,256
∑X		1,297	25,283	60,483			87,165	83,415	148,455

Table 2: The word relatedness distribution in the 'gold-standard' across each matrix

We select 100 seed nouns and 100 seed verbs with term frequencies of around 10,000 times in BNC. The average frequency of these nouns is about 8,988.9, and 10,364.4 for these verbs. High frequency words are likely to be generic or general terms and the less frequent words may not happen in the semantic sets. The average frequency of the nouns in **An**$_X$, **aN**$_X$, **Sv**$_X$, and **vO**$_X$ is in fact decreased to 3,361.1, 5,629.1, 1,156.7, and 1,692.1, and the verbs in **rV**$_X$, **Vo**$_X$, and **sV**$_X$ are decreased to 3,014.3, 3,328.9, and 1,971.8, as we only extracted syntactic dependencies from BNC. Overall, the average frequency of the nouns is about 2,959.7 across **An**$_X$, **aN**$_X$, **Sv**$_X$, and **vO**$_X$, and 3,960.9 for the verbs across **rV**$_X$, **Vo**$_X$, and **sV**$_X$.

We first used Sim_{WN} and Sim_{RT} to compare each seed word to all other words from the dependency sets, namely **An**$_X$, **aN**$_X$, **Sv**$_X$, and **vO**$_X$ for nouns and **rV**$_X$, **Vo**$_X$, and **sV**$_X$ for verbs, to retrieve its candidate words in the 'gold standard'. Instead of a normal thesaurus with a full coverage of PoS tags, we only compiled the synonyms of nouns and verbs that account for the major part of published thesauri and are more informative than other PoS tags. The word distribution within different distances to the 100 nouns and 100 verbs in the 'gold-standard' are listed in Table 2, where ∑X indicates the overall nouns from **An**$_X$, **aN**$_X$, **Sv**$_X$, and **vO**$_X$ and verbs from **rV**$_X$, **Vo**$_X$, and **sV**$_X$ in the 'gold-standard'. For the 'gold-standard' words from WordNet, SA denotes syn/antonyms of the targets, and DI the words with exactly I link distance to targets (for nouns $I \leq \gamma = 4$; for verbs $I \leq \gamma = 2$); ∑ denotes the total number of 'gold-standard' words in each matrix; and Total means the overall number of 'gold-standard' words from both WordNet and Roget. In Table 2 the average number of 'gold-standard' words across each matrix is evenly distributed.

The agreement between the WordNet-style and Roget-style words in the 'gold-standard' across these matrices, that is, the ratio of the number of words retrieved by Sim_{WN} and Sim_{RT} in both WordNet and Roget against the total number of 'gold-standard' words, is on average 7.3% on nouns and less than 15.2% on verbs. We aggregated all the 'gold-standard' words across **An**$_X$, **aN**$_X$, **Sv**$_X$, and **vO**$_X$ for nouns, as well as **rV**$_X$, **Vo**$_X$, and **sV**$_X$ for verbs, which results in 244,245 nouns and 148,455 verbs overall in the 'gold standard'. The agreement between WordNet and Roget candidates on nouns and verbs is respectively about 6.9% and 14.9%, that is to say, about 14.8% and 11.6% nouns in WordNet and Roget are of same, so are 25.4% and 26.5% for verbs. Each target noun on average owns about 1,148 WordNet, 1,464 Roget, and 2442 Total words in the 'gold standard', and each target verb 872, 834, and 1485 words respectively.

4.4 A walk-through example

For each seed word, after computing the *cosine* similarity of the seed with all other words in each dependency matrix, we produced and ranked the top n words as candidates. We then applied the two heuristics: '***any two***' and '***all***' on these candidates to forming automatic thesauri.

In Table 3 we exemplify the top 20 similar words of *sentence* and *attack* yielded in each dependency set and the two heuristics. Consider the distributionally similar words of *sentence* and *attack* in **aN**$_X$ and **rV**$_X$ for example. The words related to the linguistic sense of *sentence* consists of *syllable*, *words*, *adjective*, etc, in **aN**$_X$**,** while the words with the judicial sense make up around half of the 20 words including *imprisonment*, *penalty*, and the like. The words such as *rape* and *slaughter* from **rV**$_X$ are from the literal sense of *attack*, together with its metaphorical sense among other words like *badmouth*, *flame*, and so on.

The heuristic of '***any two***' collected the intersection of thesaurus items across these dependency sets. For example, *punishment* and *words* are the similar words to *sentence*, which respectively occurred in **aN**$_X$ and **vO**$_X$ as well as in **aN**$_X$ and **An**$_X$; *criticise* and *bomb* are the similar words to *attack*, which respectively occurred in **Vo**$_X$ and **rV**$_X$ as well as in **Vo**$_X$ and **sV**$_X$.

	Similar words
aN_X	*imprisonment term utterance penalty excommunication syllable words punishment prison prisoner phrase detention hospitalisation fisticuffs banishment verdict Minnesota meaning adjective warder*
An_X	*words syllable utterance clause nictation word swarthiness paragraph text homograph discourse imprisonment nonce phrase hexagram adjective verb niacin savarin micheas*
vO_X	*soubise cybele sextet cristal raper stint concatenation kohlrabi tostada apprenticeship ban contrivance Guadalcanal necropolis misanthropy roulade gasworks curacy jejunum punishment*
Sv_X	*ratel occurrence cragsman jingoism shiism Oklahoma genuineness unimportance language gathering letting grimm chaucer accent taxation ultimatum arrogance test verticality habituation*
any two	*imprisonment words utterance word term punishment paragraph text phrase jail verb meaning noun poem language passage sequence syllable lexicon fine*
all	*Imprisonment utterance penalty excommunication punishment prison prisoner detention hospitalisation banishment Minnesota meaning contrariety phoneme consonant counterintelligence starvation fine cathedra lifespan*

(a) The similar words to *sentence* (as a noun)

	Similar words
rV_X	*assault rape criticize arm slaughter abduct mortar accuse defend fire avow lash badmouth blaspheme slit singe flame kidnap persecute*
Vo_X	*Raid criticise bomb realign outwit beleaguer guard raze bombard criticize resemble spy pulse misspend reformulate alkalinise metastasise placard ruck glory*
sV_X	*ambush invade fraternize palpitate patrol wound pillage bomb billet shell fire liberate kidnap raid garrison accuse assault arrest slaughter outnumber*
any two	*assault criticize bomb ambush accuse raid fire rape bombard kidnap infiltrate patrol defend storm invade arrest garrison torture stab shoot*
all	*raid bomb assault criticize ambush accuse fire guard bombard patrol rape storm infiltrate wound kidnap criticise garrison alkalinize torture spy*

(b) The similar words to *attack* (as a verb)

Table 3: A sample of thesaurus items

4.5 Performance evaluation

Instead of simply matching with the 'gold standard' thesauri, Lin (1998) proposed to compare his automatic thesaurus with WordNet and Roget on their structures, taking into account the similarity scores and orders of similar words respectively produced from distributional similarity and taxonomic similarity. This approach can account for thesaurus resemblance under the hierarchy of WordNet or Roget, which is an apparent advantage over straight word matching.

Instead of calculating the varied cosine similarity between each target vector yielded from automatic thesaurus and from WordNet or Roget (Lin, 1998), we adapted the concept of Precision (*Pn*) and Recall-precision (*Rp*) from information retrieval to demonstrate much sensible values of precision and recall for a ranked list. Given the top *n* similar words *S* for a target *T* in an automatic thesaurus *Pn* is defined as $|S|/n$, where $|S|$ refers to the number of *S* that can be retrieved in the top *n* similar words of *T* in WordNet or Roget. *Rp* is conditioned on precision and is correspondingly defined as $|S|/\sum d(S)$, where in terms of words $d(S)$ denotes minimum distance between *T* and *S* if *S* can be located within the top *n* similar words of *T* in WordNet or Roget. Analogously for the ranked word list from an automatic thesaurus, the top *n* similar words with respect to each sense of *T* in WordNet are produced in the order of hyper/hyponyms and holo/meronyms with exhausting initially synonyms and then antonyms, whereas the top *n* words in Roget can be subsequently acquired within +/-*n* (preceding/succeeding) words from *T* in each of its category. Through these redefined precision and recall *Pn* can stand for the coverage of the automatic thesaurus on potentially arbitrary senses or categories of T and *Rp* can describe relatedness of the thesaurus on the actual sense or category of *T*.

5 Results

We took the top *n* similar words derived from each co-occurrence matrix for '***any two***' or '***all***', with *n* varying from 1 to 1000 in ten steps, roughly doubling each time. The results are shown in Table 4. We individually listed *Pn* and *Rp* values with respect to WordNet, Roget, and the union of WordNet and Roget (Total).

		'*all*'						'*any two*'					
		WordNet		Roget		Total		WordNet		Roget		Total	
N		*Pn*	*Rp*	*Pn*	*Rp*	*Pn*	*Rp*	*Pn*	*Rp*	*Pn*	*Rp*	*Pn*	*Rp*
1	noun	22.0	22.0	15.0	15.0	27.0	27.0	24.0	24.0	12.0	12.0	**28.0**	**28.0**
	verb	13.0	13.0	7.0	7.0	16.0	16.0	15.0	15.0	8.0	8.0	**20.0**	**20.0**
2	noun	31.0	35.2	19.0	23.7	36.0	**41.2**	34.0	34.0	20.0	20.0	**42.0**	37.5
	verb	39.0	31.7	9.5	12.0	40.0	34.2	48.5	34.4	11.0	13.3	**49.5**	**38.2**
5	noun	42.4	21.1	22.2	29.5	46.8	**27.1**	56.6	17.1	28.4	24.0	**63.2**	20.0
	verb	54.2	25.6	20.2	17.1	55.8	26.9	62.6	27.4	23.8	15.0	**64.0**	**28.7**
10	noun	43.4	11.8	19.4	18.5	47.5	**15.5**	56.6	10.4	26.9	17.1	**62.3**	11.0
	verb	53.3	19.5	18.0	17.5	54.7	19.6	62.3	21.7	20.9	15.9	**63.7**	**21.2**
20	noun	37.7	9.5	16.1	13.8	41.6	**9.8**	50.2	8.7	22.7	16.5	**56.0**	8.4
	verb	49.3	15.0	13.9	15.0	50.9	14.7	57.5	15.6	16.1	13.8	**59.0**	**15.4**
50	noun	29.0	8.0	11.2	11.2	32.3	**7.4**	41.4	7.2	16.7	9.5	**46.4**	6.8
	verb	43.8	11.9	10.0	10.9	45.4	11.3	49.5	12.2	11.4	9.9	**51.3**	**11.5**
100	noun	22.9	8.4	8.2	9.5	25.7	**7.4**	33.8	6.6	12.8	6.6	**38.4**	5.9
	verb	39.7	10.0	7.7	8.4	41.2	9.2	44.1	10.4	8.4	7.5	**45.6**	**9.8**
200	noun	18.6	6.9	5.9	7.8	20.9	**5.9**	26.6	6.2	8.9	6.2	**30.2**	5.5
	verb	36.0	9.3	5.9	6.5	37.4	**8.6**	39.6	9.3	6.4	6.2	**41.0**	8.5
500	noun	13.6	6.4	3.9	6.1	15.4	**5.5**	18.6	6.0	5.4	5.8	**21.0**	5.3
	verb	32.6	8.5	4.2	5.7	33.8	**7.7**	35.1	8.5	4.6	5.3	**36.4**	**7.7**
1000	noun	11.0	6.3	2.8	5.5	12.4	**5.4**	14.1	6.1	3.6	5.5	**16.0**	5.2
	verb	30.5	8.2	3.4	4.9	31.6	**7.3**	32.7	8.2	3.6	4.9	**33.8**	**7.3**

Table 4: The *precision* and *recall* in automatic thesauri under the heuristics of '*any two*' and '*all*' (percentage)

6 Discussion

6.1 '*any two*' vs '*all*'

It is clear that in terms of *Pn* measurement '***any two***' consistently outperformed '***all***' for both nouns and verbs in thesaurus construction. The improvement in the precision of the '***any two***' clusters over the '***all***' heuristic was significant ($p < 0.05$, paired t test). This is achieved under the condition of comparable *Rp*. Before reaching the threshold 200, the overall *Rp* for verbs for '***any two***' almost stay higher than for '***all***', which is contrary in the case of nouns. Since then no noticeable difference can be observed. The reason behind this could be that some 'gold-standard' words derived from a matrix may never occur in the thesaurus entries from another matrix, which are neglected in '***any two***'.

We also extend this work to the words with intermediate (around 4,000) and low (around 1,000) term frequencies in BNC. For the 100 nouns and 100 verbs with the intermediate frequencies, 3,753.9 and 3,675.2 respectively, the average frequency of the nouns across $\mathbf{An}_X$, $\mathbf{aN}_X$, $\mathbf{Sv}_X$, and $\mathbf{vO}_X$ is 1,274.7, and the verbs across $\mathbf{rV}_X$, $\mathbf{Vo}_X$, and $\mathbf{sV}_X$ is 1,422.0. For the 100 nouns and 100 verbs with low frequencies: 824.1 and 864.6, the average frequency of the nouns across $\mathbf{An}_X$, $\mathbf{aN}_X$, $\mathbf{Sv}_X$, and $\mathbf{vO}_X$ is 297.0, and the verbs 342.2 across $\mathbf{rV}_X$, $\mathbf{Vo}_X$, and $\mathbf{sV}_X$. For the intermediate and low frequency words, the heuristic of '***any two***' still significantly outperformed the '***all***' in yielding automatic thesauri ($p < 0.05$) with higher precision.

As the threshold increasing from 1 to 1000 in Table 4, both the nominal and verbal parts of thesaurus using the heuristics of '***any two***' and '***all***' could corroborate a preference for relationships from WordNet rather than from Roget, since both *Pn* in WordNet contributed majority of the overall *Pn* in contrast to it in Roget. Note that from the figures shown in Table 2, we can observe that the overlap between WordNet and Roget is rather small, where only 14.8% of WordNet or 11.6% of Roget for nouns co-occur, so does 25.4% of WordNet or 26.5% of Roget for verbs. This could be caused by filtering out more Roget words present in the '***all***' or '***any two***' thesaurus. This trend keep unchanged even when more unrelated words could be introduced as the threshold approached 1000.

We can compare the entry of *sentence* and *attack* with the threshold of 20 in the '***any two***' thesaurus to their respective entries in the '***all***' thesaurus, that are listed in Table 3. The entry of *sentence* in the '***any two***' thesaurus constituted the top 20 similar words in Table 3 (a), they were all akin to *sentence* without any 'noisy' words such as *Minnesota* and *counterintelligence* in the '***all***' thesaurus. So did *attack* in Table 3 (b), which comprised near-synonyms after filtering out the unrelated words such as *alkalinise* in the '***all***' thesaurus. However, some truly related words were also missed out in the '***any two***' thesauri, for example, the similar words *penalty* and *banishment* to *sentence* in the '***all***' thesaurus, as well as *guard* and *wound* to *attack*. This can be partly complemented through increasing the threshold. Even with the threshold 50, the overall thesaurus entries of were still acceptable with approximately 50% of total precision.

6.2 The predominant sense

Word senses in WordNet are ranked by their frequencies, where the first sense often serves as the predominant sense of a word. The predominant sense often serves as a back-off in sense disambiguation. To study the sense distribution of the words in automatic thesaurus, we also calculated *Pn* on the condition of extracting the 'gold-standard' words exclusively related to the first sense of a target (*First*), in contrast to all the senses.

Overall the precision of *First* sense is not less than 50% of the precision of all sense for both nouns and verbs in the '***any two***' heuristic. This implies that distributionally similar words derived using the '***any two***' heuristic are more semantically related to the first sense of a target, around 50% or more, than other senses. Even in the '***all***' heuristic around 50% of the words that match a 'gold-standard' for any sense, hold semantic relatedness with the first senses of targets.

The unbalanced sense distribution among the thesaurus items shows the uneven usages of words with respect to the Zipf's Law (1965). Kilgarriff (2004) also noted Zipfian distribution of both word sense and words when analysing the Brown corpus and BNC. The predominant sense of a word can be formed through their distributionally similar words instead of laborious sense annotation work, which serves as an important resource in sense disambiguation.

6.3 Distributional similarity and semantic relatedness

Semantic similarity is often regarded as a special case of semantic relatedness, while the latter also contains word association. Distributional similarity consists of both semantic similarity and word association between a seed word and candidate words in its thesaurus items, except for the 'noisy' words (due to the parsing or statistical errors) that hold no plausible relationships with the seed. Consider the distributionally similar words of *sentence* produced in $\mathbf{aN}_X$ in Table 3 (a) for example. Only three words, namely *term*, *phrase*, and *verdict*, were connected with *sentence* through the similarity measurement of Sim_{WN} in WordNet, whereas 14 words such as *phrase* and *penalty* shared the same topics with *sentence* in Roget. The noun *sentence* consists of three senses in WordNet,

- *sentence#n#*1: a string of words satisfying the grammatical rules of a language
- *sentence#n#*2: (criminal law) a final judgment of guilty in a criminal case and the punishment that is imposed
- *sentence#n#*3: the period of time a prisoner is imprisoned

The word *sentence* is also located in Section 480 (*Judgement*), 496 (*Maxim*), 535 (*Affirmation*), 566 (*Phrase*), and 971 (*Condemnation*) in Roget. For example, the nominal part of Section 480 is,

480. Judgment. [Conclusion.]

N. result, conclusion, upshot; deduction, inference, ergotism[Med]; illation; corollary, porism[obs3]; moral. estimation, valuation, appreciation, judication[obs3]; dijudication[obs3], adjudication; arbitrament, arbitrement[obs3], arbitration;

*assessment, ponderation[obs3]; valorization. award, estimate; review, criticism, critique, notice, report. decision, determination, judgment, finding, verdict, **sentence**, decree; findings of fact; findings of law; res judicata[Lat]. plebiscite, voice, casting vote; vote &c. (choice) 609; opinion &c. (belief) 484; good judgment &c. (wisdom) 498. judge, umpire; arbiter, arbitrator; assessor, referee. censor, reviewer, critic;*

connoisseur; commentator &c. 524; inspector, inspecting officer. twenty-twenty hindsight [judgment after the fact]; armchair general, Monday morning quarterback.

Generally *sentence#n#*1 in WordNet can be projected into Section 496 and 566, and *sentence#n#*2 into Section 480 and 971, and *sentence#n#*3 into Section 535. With respect to the evaluation of Sim_{WN} in WordNet, *term* in Table 3 (a) is the hypernym of *sentence#n#*3; and *phrase* and *sentence#n#*1 distance themselves in three links, say, *sentence#n#*1 has a meronym of *clause* that is a coordinate of *phrase*; and *sentence#n#*2 bears the same hypernym with *verdict* within four links. Apart from the paradigmatic relationships in WordNet, the three words also connect with *sentence* through Sim_{RT} in Roget, where words such as *verdict* and *sentence*s are located under the same section—*Judgement* (480). However, *sentence* holds more relations of being in the same domain with its similar words in the thesaurus from $\mathbf{aN}_{\mathrm{X}}$. For example, *penalty* and *sentence* come from/exist in Section 971, which expresses the notion of criminality deserving a penalty in a way of judicial sentence, and *prisoner* and *sentence* are situated in Section 971, which illustrates being in prison resulting from judgements in a court in the context of criminal law.

As we compute distributional similarity on the assumption of similar words sharing similar contexts conditioned by grammatical relations, in general more paradigmatic relations can be found than syntagmatic ones. In Table 4, the higher precision for WordNet than for Roget's Thesaurus show that distributionally similar words are more semantically similar rather than associated words. This is consistent with the conclusion of Kilgarriff and Yallop (2000) on computing distributional similarity that the hypothesis of similar words sharing similar contexts constrained by grammatical relations can yield *tighter* or WordNet-style thesauri, whereas the hypothesis of similar words sharing unconditioned co-occurrences can yield *looser* or Roget-style thesauri. Note that distributionally similar words could be semantically opposite to each other, given the common grammatical relations they often share. For example, in the automatic thesaurus produced with '***any two***', the nouns *failure* and *success*, or *strength* and *weakness*, are antonymous, as well the verbs *cry* and *laugh*, *deny* and *admit*.

It is clear that the 'gold standard' is subject to the vocabulary size of WordNet and Roget's Thesaurus. The worse case is from the 1911 version of Roget's Thesaurus we adopted, where words generated in modern times are not contained. For example words such as *software* and its distributionally similar words, including *emulator*, *unix*, *NT*, *Cobol*, *Oracle* (as the database system), *processor*, and *PC*, are not included in the 1911 version of Roget. We selected the target word with relatively higher frequencies in BNC and did a simple morphology analysis in the construction of the matrices using word-mapping table in WordNet, so that all nouns and verbs from automatic term clustering can be covered (at least in WordNet). However, not all word relationships in automatic thesauri could be contained in WordNet, even though we have included Roget to supply richer relationships. For example, take the words *sentence* and *detention.* In Table 3 (a) *detention* is listed in the top 20 similar words to *sentence* on $\mathbf{aN}_{\mathrm{X}}$, but they have no direct or indirect links in WordNet, nor are they situated under any *topic* or *section* in Roget, but their intense association has become commonly used. Likewise, *kidnap* as one of the top 20 similar words to *attack* on $\mathbf{rV_X}$ in Table 3 (b), which is distributionally similar to *attack,* but there are no existing connections between them in WordNet and Roget.

7 Conclusion

With the introduction of grammatical relations in computing distributional similarity, automatic thesaurus construction can be improved through the interchangeability of similar words in diverse syntactically conditional contexts. Most methods still combined these contexts into one united representation for similarity computation, which worked analogously to these based on the premise of '*a bag of words*'. After the categorization of the syntactically conditioned contexts, through which similar words can be formed under the assumption of context interchangeability, automatic thesauri were yielded with significantly higher precision than the traditional methods. Future research will focus on clustering dependencies and extracting word senses from the thesaurus entries. Learning or enriching ontologies from automatic thesauri is also the next task.

8 References

Carroll, John, Ted Briscoe and Antonio Sanfilippo (1998). Parser Evaluation: a Survey and a New Proposal. In the First International Conference on Language Resources and Evaluation, 447-454. Granada, Spain.

Cruse, D. A. (1986). Lexical Semantics, Cambridge University Press.

Curran, James R. (2003). From Distributional to Semantic Similarity. Ph.D thesis

Deerwester, Scott C., Susan T. Dumais, Thomas K. Landauer, George W. Furnas and Richard A. Harshman (1990). Indexing by Latent Semantic Analysis. Journal of the American Society of Information Science 41(6): 391-407.

Fellbaum, Christiane (1998). WordNet: An Electronic Lexical Database. Cambridge, MA, The MIT Press.

Grefenstette, Gregory (1992). Sextant: Exploring Unexplored Contexts for Semantic Extraction from

Syntactic Analysis. In the 30th Annual Meeting of the Association for Computational Linguistics, 324-326. Newark, Delaware.

Grefenstette, Gregory (1993). Evaluation Techniques for Automatic Semantic Extraction: Comparing Syntactic and Window Based Approaches. In the Workshop on Acquisition of Lexical Knowledge from Text, 143-153.

Harris, Zellig (1985). Distributional Structure. In The Philosophy of Linguistics J. J. Katz, (ed). New York, Oxford University Press: 26-47.

Hindle, Donald (1990). Noun Classification from Predicate-argument Structures. In the 28th Annual Meeting of the Association for Computational Linguistics, 268-275. Pittsburgh, Pennsylvania.

Hirschman, Lynette, Ralph Grishman and Naomi Sager (1975). Grammatically-based Automatic Word Class Formation. Information Processing and Management 11: 39-57.

Kilgarriff, Adam (1997). I don't Believe in Word Senses. Computers and the Humanities 31(2): 91-113.

Kilgarriff, Adam (2004). How Dominant is the Commonest Sense of a Word? In the 7th International Conference (TSD 2004, Text, Speech and Dialogue), 103-112. Brno, Czech Republic.

Kilgarriff, Adam and Colin Yallop (2000). What's in a Thesaurus? In the Second International Conference on Language Resources and Evaluation, LREC-2000, 1371-1379. Athens, Greece.

Landauer, Thomas K. and Susan T. Dumais (1997). A Solution to Plato's Problem: the Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge. Psychological Review 104: 211-240.

Lin, Dekang (1997). Using Syntactic Dependency as a Local Context to Resolve Word Sense Ambiguity. In the 35th Annual Meeting of the Association for Computational Linguistics, 64-71. Madrid, Spain.

Lin, Dekang (1998). Automatic Retrieval and Clustering of Similar Words. In the 17th International Conference on Computational Linguistics, 768-774. Montreal, Quebec, Canada.

Lowe, Will (2001). Towards a Theory of Semantic Space. In the 23rd Annual Conference of the Cognitive Science Society, 576-581. Edinburgh, UK.

McCarthy, Diana, Rob Koeling, Julie Weeds and John Carroll (2004). Finding Predominant Senses in Untagged Text. In the 42nd Annual Meeting of the Association for Computational Linguistics (ACL-04), 267-287. Barcelona, Spain.

Miller, George A. and Walter G. Charles (1991). Contextual Correlates of Semantic Similarity. Language and Cognitive Processes 6(1): 1-28.

Molla, Diego and Ben Hutchinson (2003). Intrinsic versus Extrinsic Evaluations of Parsing Systems. In European Association for Computational Linguistics(EACL), workshop on Evaluation Initiatives in Natural Language Processing, 43-50. Budapest, Hungary.

Padó, Sebastian and Mirella Lapata (2007). Dependency-based construction of semantic space models. To appear in Computational Linguistics 33(2).

Pantel, Patrick and Dekang Lin (2002). Discovering Word Senses from Text. In the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 613-619. New York, NY, USA.

Resnik, Philip (1997). Selectional Preference and Sense Disambiguation. In ACL Siglex Workshop on Tagging Text with Lexical Semantics, Why, What and How?, 52-57. Washington, USA.

Sahlgren, Magnus (2006). The Word-Space Model: Using Distributional Analysis to Represent Syntagmatic and Paradigmatic Relations between Words in High-Dimensional Vector Spaces. Ph.D thesis

Salton, Gerard and Michael J. McGill (1986). Introduction to Modern Information Retrieval. New York, NY, USA, McGraw-Hill.

Schütze, Hinrich (1992). Dimensions of Meaning. In the 1992 ACM/IEEE Conference on Supercomputing, 787-796. Minneapolis, Minnesota, USA.

Sleator, Daniel and Davy Temperley (1991). Parsing English with a Link Grammar, Carnegie Mellon University.

Weeds, Julie Elizabeth (2003). Measures and Applications of Lexical Distributional Similarity. Ph.D thesis

Yang, Dongqiang and David M.W. Powers (2005). Measuring Semantic Similarity in the Taxonomy of WordNet. In the Twenty-Eighth Australasian Computer Science Conference (ACSC2005), 315-322. Newcastle, Australia, ACS.

Yang, Dongqiang and David M.W. Powers (2006). Verb Similarity on the Taxonomy of WordNet. In the 3rd International WordNet Conference (GWC-06), 121-128. Jeju Island, Korea.

Yarowsky, David (1993). One Sense per Collocation. In ARPA Human Language Technology Workshop, 266-271. Princeton, New Jersey.

Zipf, George Kingsley (1965). Human Behavior and the Principle of Least Effort: an Introduction to Human Ecology. N.Y., Hafner Pub. Co.