

# Word Sense Disambiguation using lexical cohesion in the context

Dongqiang Yang | David M.W. Powers  
School of Informatics and Engineering  
Flinders University of South Australia  
PO Box 2100, Adelaide

Dongqiang.Yang|David.Powers@flinders.edu.au

## Abstract

This paper designs a novel lexical hub to disambiguate word sense, using both syntagmatic and paradigmatic relations of words. It only employs the semantic network of WordNet to calculate word similarity, and the Edinburgh Association Thesaurus (EAT) to transform contextual space for computing syntagmatic and other domain relations with the target word. Without any back-off policy the result on the English lexical sample of SENSEVAL-2<sup>1</sup> shows that lexical cohesion based on edge-counting techniques is a good way of unsupervisedly disambiguating senses.

## 1 Introduction

Word Sense Disambiguation (WSD) is generally taken as an intermediate task like part-of-speech (POS) tagging in natural language processing, but it has not so far achieved the sufficient precision for application as POS tagging (for the history of WSD, cf. Ide and Véronis (1998)). It is partly due to the nature of its complexity and difficulty, and to the widespread disagreement and controversy on its necessity in language engineering, and to the representation of the senses of words, as well as to the validity of its evaluation (Kilgarriff and Palmer, 2000). However the endeavour to automatically achieve WSD has been continuous since the earliest work of the 1950's.

In this paper we specifically investigate the role of semantic hierarchies of lexical knowledge on WSD, using datasets and evaluation methods from SENSEVAL (Kilgarriff and Rosenzweig,

2000) as these are well known and accepted in the community of computational linguistics.

With respect to whether or not they employ the training materials provided, SENSEVAL roughly categorizes the participating systems into “unsupervised systems” and “supervised systems”. Those that don't use the training data are not usually truly unsupervised, being based on lexical knowledge bases such as dictionaries, thesauri or semantic nets to discriminate word senses; conversely the “supervised” systems learn from corpora marked up with word senses.

The fundamental assumption, in our “unsupervised” technique for WSD in this paper, is that the similarity of contextual features of the target with the pre-defined features of its sense in the lexical knowledge base provides a quantitative cue for identifying the true sense of the target.

The lexical ambiguity of polysemy and homonymy, whose distinction is however not absolute as sometimes the senses of word may be intermediate, is the main object of WSD. Verbs, with their more flexible roles in a sentence, tend to be more polysemous than nouns, so worsening the computational feasibility. In this paper we disambiguated the sense of a word after its POS tagging has assigned them either a noun or a verb tag. Furthermore, we deal with nouns and verbs separately.

## 2 Some previous work on WSD using semantic similarity

Sussna (1993) utilized the semantic network of nouns in WordNet to disambiguate term senses to improve the precision of SMART information retrieval at the stage of indexing, in which he assigned two different weights for both directions of edges in the network to compute the similarity of two nodes. He then exploited the moving fixed size window to minimize the sum

---

<sup>1</sup> <http://www.senseval.org/>

of all combinations of the shortest distances among target and context words.

Pedersen et al. (2003) extended Lesk's definition method (1986) to discriminate word sense through the definitions of both target and its IS-A relatives, and achieved a better result in the English lexical sample task of SENSEVAL-2, compared with other edge-counting or statistical estimation metrics on WordNet.

Humans carefully select words in a sentence to express harmony or cohesion in order to ease the ambiguity of the sentence. Halliday and Hasan (1976) argued that cohesive chains unite text structure together through reiteration of reference and lexical semantic relations (superordinate and subordinate). Morris and Hirst (1991) suggested building lexical chains is important in the resolution of lexical ambiguity and the determination of coherence and discourse structure. They argued that lexical chains, which cover the multiple semantic relations (systematic and non-systematic), can transform the context setting into the computational one to narrow down the specific meaning of the target, manually realizing this with the help of Roget's Thesaurus. They defined a lexical chain within Roget's very general hierarchy, in which lexical relationships are traced through a common category.

Hirst and St-Onge (1997) define a lexical chain using the syn/antonym and hyper/hyponym links of WordNet to detect and correct malapropisms in context, in which they specified three different weights from extra-strong to medium strong to score word similarity to decide the inserting sequence in the lexical chain. They first computationally employed WordNet to form a "greedy" lexical chain as a substitute of the context to solve the matter of malapropism, where the word sense is decided by its preceding words.

Around the same time, Barzilay and Elhadad (1997) realized a "non-greedy" lexical chain, which determined the word sense after processing of all words, in the context of text summarization.

In this paper we propose an improved lexical chain, the *lexical hub*, that holds the target to be disambiguated as the centre, replacing the usual chain topology used in text summarization and cohesion analysis. In contrast with previous methods we only record the lexical hub of each sense of the target, and we don't keep track of other context words. In other words, after the computation of lexical hub of the target, we can immediately produce the right sense of the target even though the senses of the context words are

still in question. We also transform the context surroundings through a word association thesaurus to explore the effect of other semantic relationships such as syntagmatic relation against WSD.

### 3 Selection of knowledge bases

WordNet (Fellbaum, 1998) provides a fine-grained enumerative semantic net that is commonly used to tag the instances of English target words in the tasks of SENSEVAL with different senses (WordNet synset numbers). WordNet groups related concepts into synsets and links them through IS-A and PART-OF links, emphasizing the vertical interaction between the concepts that is much paradigmatic.

Although WordNet can capture the fine-grained paradigmatic relations of words, another typical word relationship, syntagmatic connectedness, is neglected. The syntagmatic relationship, which is often characterized with different POS tag, and frequently occurs in corpora or human brains, plays a critical part in cross-connecting words from different domains or POS tags.

It should be noted that WordNet 2.0 makes some efforts to interrelate nouns and verbs using their derived lexical forms, placing associated words under the same domain. Although some verbs have derived noun forms that can be mapped onto the noun taxonomy, this mapping only relates the morphological forms of verbs, and still lacks syntagmatic links between words.

The interrelationship of noun and verb hierarchies is far from complete and only a supplement to the primary IS-A and PART-OF taxonomies in WordNet. Moreover as WordNet generally concerns the paradigmatic relations (Fellbaum, 1998), we have to seek for other lexical knowledge sources to compensate for the shortcomings of WordNet in WSD.

The Edinburgh Association Thesaurus<sup>2</sup> (EAT) provides an associative network to account for word relationship in human cognition after collecting the first response words for the stimulus words list (Kiss et al., 1973). Take the words *eat* and *food* for example. There is no direct path between the concepts of these two words in the taxonomy of WordNet (both as noun and verb), except in the gloss of the first and third sense of *eat* to explain 'take in solid food', or 'take in food', which glosses are not regularly or care-

---

<sup>2</sup> <http://www.eat.rl.ac.uk/>

fully organized in WordNet. However in EAT *eat* is strongly associated with *food*, and when taking *eat* as a stimulus word, 45 out of 100 subjects regarded *food* as the first response.

Yarowsky (1993) indicated that the objects of verbs play a more dominant role than their subjects in WSD and nouns acquire more stable disambiguating information from their noun or adjective modifiers.

In the case of verbs association tests, it is also reported that more than half the response words of verbs (the stimuli) are syntagmatically related (Fellbaum, 1998). In experiments of examining the psychological plausibility of WordNet relationships, Chaffin et al. (1994) stated that only 30.4% of the responses of 75 verb stimuli belongs to verbs, and more than half of the responses are nouns, of which nearly 90% are categorized as the arguments of the verbs.

Sinopalnikova (2004) also reported that there are multiple relationships found in word association thesaurus, such as syntagmatic, paradigmatic relations, domain information etc.

In this paper we only use the straightforward forms of context words separating the effect of syntactic dependence on the WSD. As a supplement of enriching word linkage in the WSD, we retrieve the lexical knowledge from both WordNet and EAT. We first explore the function of semantic hierarchies of WordNet on WSD, and then we transform the context word with EAT to investigate whether other relationships can improve WSD.

## 4 System design

In order to find semantically related words to cohesively form lexical hubs, we first employ the two word similarity algorithms of Yang and Powers (2005; 2006) that use WordNet to compute noun similarity and verb similarity respectively. We next construct the lexical hub for each target sense to assemble the similarity score between the target and its context words together. The maximum score of these lexical hubs specifically predicts the real sense of the target, also implicitly captures the cohesion and real meaning of the word in its context.

### 4.1 Similarity metrics on nouns

Yang and Powers (2005) designed a metric,

$$Sim(c1, c2) = \alpha_i * \beta^{\lambda}$$

utilizing both IS-A and PART-OF taxonomies of WordNet to measure noun similarity, and they argued that the similarity of nouns is the maxi-

imum of all their concept similarities. They defined the similarity (*Sim*) of two concepts (*c1* and *c2*) with a link type factor ( $\alpha_t$ ) to specify the weights of different link types (*t*) (syn/antonym, hyper/ hyponym, and holo/meronym) in the WordNet, and a path type factor ( $\beta_t$ ) to reduce the uniform distance of the single link, along with a depth factor ( $\lambda$ ) to restrict the maximum searching distance between concepts. Since their metric on noun similarity is significantly better than some popular measures and even outperforms some subjects on a standard data set, we selected it as a measure on noun similarity in our WSD task.

### 4.2 Similarity metrics on verbs

Yang and Powers (2006) also redesigned their noun model,

$$Sim(c1, c2) = \alpha_{str} * \alpha_t * \prod_{i=1}^{Dist(c1, c2)} \beta_i$$

to accommodate verb case, which is harder to deal with in the shallow and incomplete taxonomy of verbs in WordNet. As an enhancement to the uniqueness of verb similarity they also consider three fall-back factors, where if  $\alpha_{str}$  is 1 normally but successively falls back to:

- $\alpha_{stm}$ : the verb stem polysemy ignoring sense and form
- $\alpha_{der}$ : the cognate noun hierarchy of the verb
- $\alpha_{gls}$ : the definition of the verb

They also defined two alternate search protocols: rich hierarchy exploration (RHE) with no more than six links and shallow hierarchy exploration (SHE) with no more than two links.

One minor improvement to the verb model in their system comes from comparing the similarity of verbs and nouns using the noun model metric for the derived noun form of verb. It thus allows us to compare nouns and verbs and avoids the limitation of having to have the same POS tag.

### 4.3 Depth in WordNet

Yang and Powers fine-tuned the parameters of the noun and verb similarity models, finding them relatively insensitive to the precise values, and we have elected to use their recommended values for the WSD task. But it is worth mentioning that their optimal models are achieved in purely verbal data sets, i.e. the similarity score is context-free.

In their models, the depth in the WordNet, i.e. the distance between the synsets of words ( $\lambda$ ), is indeed an outside factor which confines the searching scope to the cost of computation and depends on the different applications. If we tuned it using the training data set of SENSEVAL-2 we probably would assign different values and might achieve better results. Note that for both nouns and verbs we employ RHE (rich hierarchy exploration) with  $\lambda = 2$  making full use of the taxonomy of WordNet and making no use of glosses.

#### 4.4 How to setup the selection standard for the senses

Other than making the most of WSD results, our main motive for this paper is to explore to what extent the semantic relationships will reach accuracy, and to fully acknowledge the contribution of this single attribute working on WSD, which is encouraged by SENSEVAL in order to gain further benefits in this field (Kilgarriff and Palmer, 2000). Without any definition, which is previously surveyed by Lesk (1986) and Pedersen et al. (2003), we screen off the definition factor in the metric of verb similarity, with the intention of focusing on the taxonomies of WordNet.

Assuming that the lexical hub for the right sense would maximize the cohesion with other words in the discourse, we design six different strategies to calculate the lexical hub in its unordered contextual surroundings.

We first put forward three metrics to measure up the similarity of the senses of the target and the context word:

- **The maximized sense similarity**

$$Sim_{max}(T_k, C_i) = \max_j (Sim(T_k, C_{i,j}))$$

where  $T$  denotes the target,  $T_k$  is the  $k$ th sense of the target;  $C_i$  is the  $i$ th context word in a fixed window size around the target,  $C_{i,j}$  the  $j$ th sense of  $C_i$ . Note that  $T$  and  $C$  can be any noun and verb, along with  $Sim$  the metrics of Yang and Powers.

- **The average of sense similarity**

$$Sim_{ave}(T_k, C_i) = \frac{\sum_{j=1}^m Sim(T_k, C_{i,j})}{\sum_{j=1}^m Links(T_k, C_{i,j})}$$

where  $Links(T_k, C_{i,j})=1$ , if  $Sim(T_k, C_{i,j})>0$ , otherwise 0.

- **The sum of sense similarity**

$$Sim_{sum}(T_k, C_i) = \sum_{j=1}^m Sim(T_k, C_{i,j})$$

where  $m$  is the total sense number of  $C_i$ .

Subsequently we can define six distinctive heuristics to score the lexical hub in the following parts:

- **Heuristic 1 – Sense Norm (HSN)**

$$Sense(T) = \arg \max_k \left( \frac{\sum_{i=1}^l Sim_{max}(T_k, C_i)}{\sum_{i=1}^l Linkw(T_k, C_i)} \right)$$

where  $Linkw(T_k)=1$  if  $Sim_{max}(T_k, C_i)>0$ , otherwise 0

- **Heuristic 2 – Sense Max (HSM)**

An unnormalized version of HSN is:

$$Sense(T) = \arg \max_k \left( \sum_{i=1}^l Sim_{max}(T_k, C_i) \right)$$

- **Heuristic 3 – Sense Ave (HSA)**

Taking into account all of the links between the target and its context word, the correct sense of the target is:

$$Sense(T) = \arg \max_k \left( \sum_{i=1}^l Sim_{ave}(T_k, C_i) \right)$$

- **Heuristic 4 – Sense Sum (HSS)**

The unnormalized version of HSA is:

$$Sense(T) = \arg \max_k \left( \sum_{i=1}^l Sim_{sum}(T_k, C_i) \right)$$

- **Heuristic 5 – Word Linkage (HWL)**

The straightforward output of the correct sense of the target in the discourse is to count the maximum number of context words whose similarity scores with the target are larger than zero:

$$Sense(T) = \arg \max_k \left( \sum_{i=1}^l Linkw(T_k, C_i) \right)$$

- **Heuristic 6 – Sense Linkage (HSL)**

No matter what kind of relations between the target and its context are, the sense of the target, which is related to the maximum counts of senses of all its context words, is scored as the right meaning:

$$Sense(T) = \arg \max_k \left( \sum_{i=1}^l \sum_{j=1}^m Links(T_k, C_{i,j}) \right)$$

Therefore the lexical hub of each sense of the target only relies on the interaction of the target and its each context word, rather than of the context words. The implication is that the lexical hub only disambiguates the real sense of the tar-

get other than the real meaning of the context word; the maximum scores or link numbers (on the level of words or senses) in the six heuristics suggest that the correct sense of the target should cohere with as many words or their senses as practicable in the discourse.

When similarity scores are ties we directly produce all of the word senses to prevent us from guessing results. Some WSD systems in SENSEVAL handle tied scores simply using the first sense (in WordNet) of the target as the real sense. It is no doubt that the skewed distribution of word senses in the corpora (the first sense often captures the dominant sense) can benefit the performance of the systems, but at the same time it mixes up the contribution of the semantic hierarchy on WSD in our system.

## 5 Results

We evaluate the six heuristics on the English lexical sample of SENSEVAL-2, in which each target word has been POS-tagged in the training part. With the absence of taxonomy of adjectives in WordNet we only extract all 29 nouns and all 29 verbs from a total of 73 lexical targets, and then we subcategorize the test dataset into 1754 noun instances and 1806 verb instances. Since the sample of SENSEVAL-2 is manually sense-tagged with the sense number of WordNet 1.7 and our metrics are based on its version 2.0, we translate the sample and answer format into 2.0 in accordance with the system output format.

Finally, we find that each noun target has 5.3 senses on average and each verb target 16.4 senses. Hence the baseline of random selection of senses is the reciprocal of each average sense number, i.e. separately 18.9 percent for nouns and 6 percent for verbs.

In addition, SENSEVAL-2 provides a scoring software with 3 levels of schemes, i.e. fine-grained, coarse-grained and mixed-grained to produce precision and recall rates to evaluate the participating systems. According to the SENSEVAL scoring system, as we always give at least one answer, the precision is identical to the recall under the separate noun and verb datasets. So we just evaluate our systems in light of accuracy. We tested the heuristics with fine-grained precision, which required the exact match of the key to each instance.

### 5.1 Context

Without any knowledge of domain, frequency and pragmatics to guess, word context is the only

way of labeling the real meaning of word. Basically a bag of context words (after morphological analyzing and filtering stop-words) or the fine-grained ones (syntactic role, selection preference etc.) can provide cues for the target. We propose to merely use a bag of words to feed into each heuristic in case of losing any valuable information in the disambiguation, and preventing from any interference of other clues except the semantic hierarchy of WordNet.

The size of the context is not a definitive factor in WSD, Yarowsky (1993) suggested the size of 3 or 4 words for the local ambiguity and 20/50 words for topic ambiguity. He also employed Roget's Thesaurus in 100 words of window to implement WSD (Yarowsky, 1992). To investigate the role of local context and topic context we vary the size of window from one word distance away to the target (left and right) until 100 words away in nouns or 60 in verbs, until there are no increases in the context of each instance.

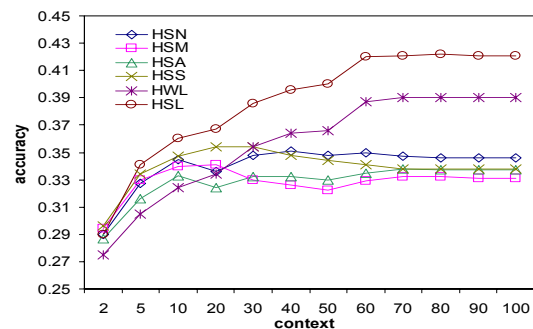


Figure 1: the result of noun disambiguation with different size of context in SENSEVAL 2

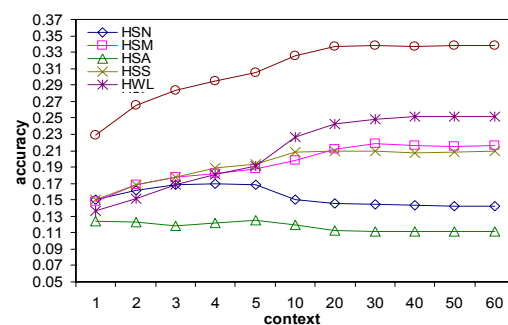


Figure 2: the result of verb disambiguation with different size of context in SENSEVAL 2

Noun and verb disambiguation results are respectively displayed in Figure 1 and 2. Since the performance curves of the heuristics turned into flat and stable (the average standard deviations of the six curves of nouns and verbs is around 0.02 level before 60 and 20, after that approxi-

mately 0.001 level), optimal performance is reached at 60 context words for nouns and 20 words for verbs. These values are used as parameters in subsequent experiments.

## 5.2 Transformed context (EAT)

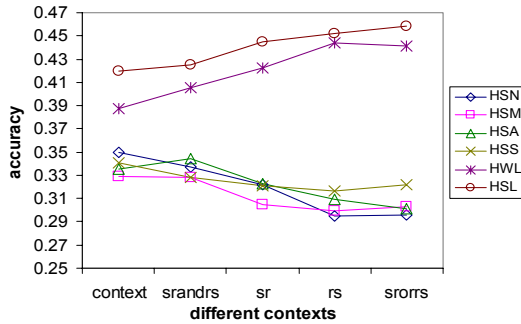


Figure 3: the results of nouns disambiguation of SENSEVAL-2 in the transformed context spaces

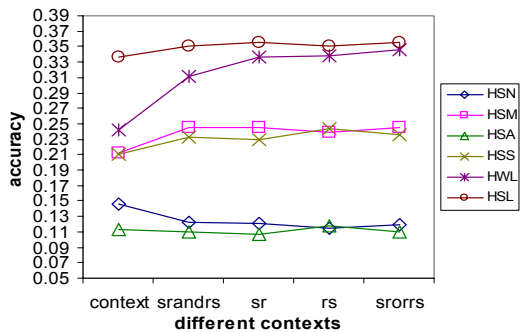


Figure 4: the results of verbs disambiguation of SENSEVAL-2 in the transformed context spaces

Although our metrics can measure the similarity of nouns and verbs through the derived related form of verbs (not from the derived verbs of nouns as a consequence of the shallowness of verb taxonomy of WordNet), we still can't completely rely on WordNet, which focuses on the paradigmatic relations of words, to fully cover the complexity of contextual happenings of words.

Since the word association norm captures both syntagmatic and pragmatic relations in words, we transform the context words of the target into its associated words, which can be retrieved in the EAT, to augment the performance of the lexical hub.

There are two word lists in the EAT: one list takes each head word as a stimulus word, and then collects and ranks all response words according to their frequency of subject consensus; the other list is in the reverse order with the response as a head word and followed by the eliciting

stimuli. We denote the stimulus/response set of word as SR, respond/stimulus as RS. Apart from that we symbolize SRANDRS as the intersection of SR and RS, along with SRORRS as the union set of SR and RS. Then for each context word we retrieve its corresponding words in each word list and calculate the similarity between the target and these words including the context words.

As a result we transform the original context space of each target into an enriched context space under the function of SR, RS, SRANDRS or SRORRS.

We take the respective 60 context words of nouns and 20 words of verbs as the reference points for the transferred context experiment, since after that the performance curves of the heuristics turned into flat and stable (the average standard deviations of the six curves of nouns and verbs is around 0.02 level before 60, after that approximately 0.001 level).

After the transformations, the noun and verb results are respectively demonstrated in Figure 3 and 4.

## 6 Comparison with other techniques.

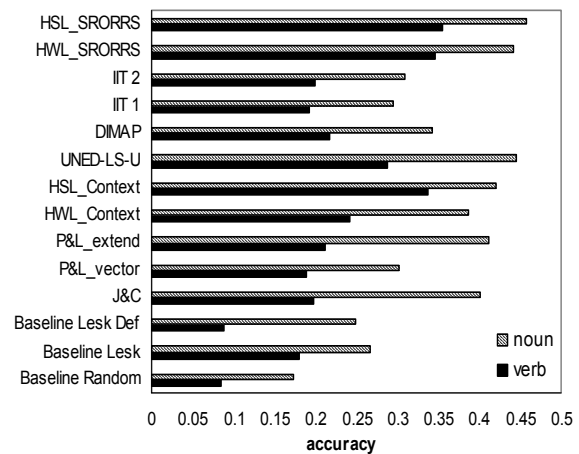


Figure 5: comparisons of HWL and HSL with other unsupervised systems and similarity metrics

Pedersen et al. (2003) in the work of evaluating different similarity techniques based on WordNet, realized two variants of Lesk's methods: extended gloss overlaps (P&L\_extend) and gloss vector (P&L\_vector), as well as evaluating them in the English lexical sample of SENSEVAL-2. The best edge-counting-based metric that they measured are from Jiang and Conrath (1997) (J&C).

Accordingly, without the transformation of EAT, we compare our results of HWL and HSL (denoted as HWL\_Context and HSL\_Context) with the above methods (picking up their optimal values). The results are illustrated in Figure 5. At the same time we also list three baselines for unsupervised systems (Kilgarriff and Rosenzweig, 2000), which are Baseline Random (randomly selecting one sense of the target), Baseline Lesk (overlapping between the examples and definitions of and unsupervised systems in SENSEVAL-2 each sense of the target and context words), and its reduced version, i.e. Baseline Lesk Def (only definition).

We further compare HWL and HSL with the intervention of SRORRS of EAT (denoted as HWL\_SRORRS and HSL\_SRORRS) with other unsupervised systems that employ no training materials of SENSEVAL-2, which are respectively:

- IIT 1 and IIT 2: extended the WordNet gloss of each sense of the target, along with its superordinate and subordinate node's glosses, without back-off policies.
- DIMAP: employed both WordNet and the New Oxford Dictionary of English. With the first sense as a back-off when tied scores occurred.
- UNED-LS-U: for each sense of the target, they enriched the sense describer through the first five hyponyms of it and a dictionary built from 3200 books from Project Gutenberg. They adopted a back-off policy to the first sense and discarded the senses accounting for less than 10 percent of files in SemCor).

## 7 Conclusion and discussion

### 7.1 Local context and topic context

On the analysis of standard deviation of precision on different stage in Figure 1 and 2 we can conclude that the optimum size for HSN to HSS was  $\pm 10$  words for nouns, reflecting a sensitivity to only local context, whilst HWL and HSL reflected significant improvement up to  $\pm 60$  reflecting a sensitivity to topical context. In the case of verbs HSA showed little significant context sensitivity, HSN showed some positive sensitivity to local context but increasing beyond  $\pm 5$  had a negative effect, HSM and HSS to HSL showed some sensitivity to broader topical context but this plateaued around  $\pm 20$  to 30.

### 7.2 The analysis of different heuristics.

HWL and HSL were clearly superior for both noun and verb tasks, with the superiority of HSL being significantly greater and more comparable between noun and verb tasks with the difference scarcely reaching significance. These observations remain true with the addition of the EAT information. After transformations with EAT for nouns, HSL and HWL no longer differ significantly in performance, forming a single group with relatively higher precision, whilst the other heuristics clump together into another group with lower precision, reflecting a negative effect from EAT. In the verb case, HWL and HSL, HSM and HSS, and HSN and HSA form three significantly different groups with reference to their precision, reflecting poor performance of both normalized heuristics (HSN and HSA) and a significantly improved result of HWL from the EAT data.

All of this implies that in the lexical hub for WSD, the correct meaning of a word should hold as many links as possible with a relatively large number of context words. These links can be in the level of word form (HWL) or word sense (HSL). HSL achieved the highest precision in both nouns and verbs.

### 7.3 The interaction of EAT in WSD

For the noun sense disambiguation, the paired two sample for mean of the t-Test showed us that RS and SRORRS transformations can significantly improve the precision of disambiguation of HWL and HSL ( $P < 0.05$ , at the confidence level of 95 percent). All four transformations using EAT for verb disambiguation are significantly better than its straightforward context case on HWL and HSL ( $P < 0.05$ , at the confidence level of 95 percent).

It demonstrated that both the syntagmatic relation and other domain information in the EAT can help discriminate word sense. With the transformation of context surroundings of the target, the similarity metrics can compare the likeness of nouns and verbs, although we can exploit the derived form of word in WordNet to facilitate the comparison.

### 7.4 Comparison with other methods

The lexical hub reached comparatively higher precision in both nouns (45.8%) and verbs (35.6%). This contrasted with other similarity based methods and the unsupervised systems in SENSEVAL-2. Note that we don't adopt any

back-off policy such as the commonest sense of word used by UNED-LS-U and DIMAP.

Although the noun and verb similarity metrics in this paper are based on edge-counting without any aid of frequency information from corpora, they performed very well in the task of WSD in relation to other information based metrics and definition matching methods. Especially in the verb case, the metric significantly outperformed other metrics.

## 8 Conclusion and future work

In this paper we defined the lexical hub and proposed its use for processing word sense disambiguation, achieving results that are comparatively better than most unsupervised systems of SENSEVAL-2 in the literature. Since WordNet only organizes the paradigmatic relations of words, unlike previous methods, which are only based on WordNet, we fed the syntagmatic relations of words from the EAT into the noun and verb similarity metrics, and significantly improved the results of WSD, given that no back-off was applied. Moreover, we only utilized the unordered raw context information without any pragmatic knowledge and syntactic information; there is still a lot of work to fuse them in the future research. In terms of the heuristics evaluated, richness of sense or word connectivity is much more important than the strength of individual word or sense linkages. An interesting question is whether these results will be borne out in other datasets. In the forthcoming work we will investigate their validity in the lexical task of SENSEVAL-3.

## References

- Barzilay, R. and M. Elhadad (1997). Using Lexical Chains for Text Summarization. In the Intelligent Scalable Text Summarization Workshop (ISTS'97), ACL, Madrid, Spain.
- Chaffin, R., et al. (1994). The Paradigmatic Organization of Verbs in the Mental Lexicon. Trenton State College.
- Fellbaum, C. (1998). Wordnet: An Electronic Lexical Database. Cambridge MA, USA, The MIT Press.
- Halliday, M. A. K. and R. Hasan (1976). Cohesion in English. London, London:Longman.
- Hirst, G. and D. St-Onge (1997). Lexical Chains as Representations of Context for the Detection and Correction of Malapropisms. Wordnet. C. Fellbaum. Cambridge, MA, The Mit Press.
- Ide, N. and J. Véronis (1998). Word Sense Disambiguation: The State of the Art. Computational linguistics 24(1).
- Jiang, J. and D. Conrath (1997). Semantic Similarity Based on Corpus Statistics and Lexical Taxonomy. In the 10th International Conference on Research in Computational Linguistics (ROCLING), Taiwan.
- Kilgarriff, A. and M. Palmer (2000). Introduction, Special Issue on Senseval: Evaluating Word Sense Disambiguation Programs. Computers and the Humanities 34(1-2): 1-13.
- Kilgarriff, A. and J. Rosenzweig (2000). Framework and Results for English Senseval. Computers and the Humanities 34(1-2): 15-48.
- Kiss, G. R., et al. (1973). The Associative Thesaurus of English and Its Computer Analysis. Edinburgh, University Press.
- Lesk, M. (1986). Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Code from an Ice Cream Cone. In the 5th annual international conference on systems documentation, ACM Press.
- Morris, J. and G. Hirst (1991). Lexical Cohesion Computed by Thesaural Relations as an Indicator of the Structure of Text. Computational linguistics 17(1).
- Pedersen, T., et al. (2003). Maximizing Semantic Relatedness to Perform Word Sense Disambiguation.
- Sinopalnikova, A. (2004). Word Association Thesaurus as a Resource for Building Wordnet. In GWC 2004.
- Sussna, M. (1993). Word Sense Disambiguation for Free-Text Indexing Using a Massive Semantic Network. In CKIM'93.
- Yang, D. and D. M. W. Powers (2005). Measuring Semantic Similarity in the Taxonomy of Wordnet. In the Twenty-Eighth Australasian Computer Science Conference (ACSC2005), Newcastle, Australia, ACS.
- Yang, D. and D. M. W. Powers (2006). Verb Similarity on the Taxonomy of Wordnet. In the 3rd International WordNet Conference (GWC-06), Jeju Island, Korea.
- Yarowsky, D. (1992). Word Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora. In the 14th International Conference on Computational Linguistics, Nates, France.
- Yarowsky, D. (1993). One Sense Per Collocation. In ARPA Human Language Technology Workshop, Princeton, New Jersey.