

Acquisition of Computational-Semantic Lexicons from Machine Readable Lexical Resources

Jason J. S. Chang and J. N. Chen
Department of Computer Science
National Tsing Hua University
Hsinchu 30043, Taiwan, ROC
Tel: +886 35 731-069 Fax: 723-694
{jschang,dr818314}@cs.nthu.edu.tw

Abstract

This paper describes a heuristic algorithm capable of automatically assigning a label to each of the senses in a machine readable dictionary (MRD) for the purpose of acquiring a computational-semantic lexicon for treatment of lexical ambiguity. Including these labels in the MRD-based lexical database offers several positive effects. The labels can be used as a coarser sense division so unnecessarily fine sense distinction can be avoided in word sense disambiguation (WSD). The algorithm is based primarily on simple word matching between an MRD definition sentence and word lists of an LLOCE topic. We also describe an implementation of the algorithm for labeling definition sentences in Longman Dictionary of Contemporary English (LDOCE). For this purpose the topics and sets of related words in Longman Lexicon of Contemporary English (LLOCE) are used in this work. Quantitative results for a 12-word test set are reported. Our discussion entails how the availability of these labels provides the means for treating such problems as: acquisition of a lexicon capable of providing broad coverage, systematic word sense shifts, lexical underspecification, and acquisition of zero-derivatives.

1. Introduction

Treatment of lexical ambiguity such as WSD has been found useful in many NLP applications, including information retrieval (McRoy 1992; Krovetz and Croft 1992) and machine translation (Brown et al. 1991; Dagan et al. 1991; Dagan and Itai 1994). Recently, various approaches (Dolan 1994; Luk 1995; Yarowsky 1992; Dagan et al. 1991; Dagan and Itai 1994) to word sense division have been used in WSD research. Directly using dictionary senses as the sense division has several advantages. First, sense distinction according to a dictionary is readily available from MRDs such as the LDOCE (Longman 1992). Second, indicative words and concepts for each sense are directly available in numbered definitions and examples. Lesk (1986) demonstrated that dictionary entries can be used to generate signatures of senses for WSD. However, using MRD as the knowledge source for sense division and disambiguation encounters certain problems. Dolan (1994) observed that sense division in MRD is frequently too fine for the purpose of WSD. A WSD system based on dictionary senses faces an unnecessary and difficult "forced-choices." Most researchers resorted to human intervention to identify and group closely related senses.

This paper describes a heuristic algorithm capable of automatically assigning a label to each of the senses in a machine readable dictionary (MRD) for the purpose of acquiring a computational-semantic lexicon for treatment of lexical ambiguity. Including these labels in the MRD-based lexical database offers several positive effects. The labels can be used as a coarser sense division so unnecessarily fine sense distinction can be avoided in word sense disambiguation (WSD). The algorithm is based primarily on simple word matching between an MRD definition sentence and word lists of an LLOCE (McArthur 1992) topic. We begin by giving the details of material used, including the characteristics of definition sentences in LDOCE and the organization of words in LLOCE. Next, the algorithm for labeling LDOCE senses is described. An illustrative example demonstrates the effectiveness of the algorithm. After describing the al-

gorithm, the experimental results for a 12-word test set are presented. Our discussion also entails the possible implication of the labels to such problems as: acquisition of a lexicon capable of providing broad coverage, systematic word sense shifts, lexical underspecification, and acquisition of zero-derivatives at the sense level. Moreover, the proposed algorithm is compared with other approaches in available literature. Finally, concluding remarks are made.

2. Identifying the topic of senses

The labeling of dictionary definition sentences with a coarse sense distinction such as the set labels in LLOCE is a special form of the WSD problem. No simple method can solve the general problem of WSD for unrestricted text. We will show that this labeling task is made simpler for several reasons. For example, consider the definition sentences for the first 5 senses of "bank" in LDOCE:

- 1 . *land* along the side of a river, lake, etc.
- 2 . *earth* which is heaped up in a field or garden, often making a border or division.
- 3.a *mass* of snow, clouds, mud, etc.
- 4.a *slope* made at bends in a road or race-track, so that they are safer for cars to go round.
- 5.= *SANDBANK* (a high underwater *bank* of sand in a river, harbour, etc.).

First of all, only simple words are used in the definitions. Furthermore, the text generation schemes are rather regular. The scheme that lexicographers used in generating the definitions above is similar to the DEFINITION scheme described in McKeown (1985). A DEFINITION scheme begins with a *genus term* (that is, conceptual parent or ancestor of the sense), followed by the so-called *differentia* that consists of words semantically related to the sense to provide specifics about the sense. Those relations between the sense and its defining words are reflected in semantic clusters that are termed *categorical*, *functional*, and *situational* clusters in McRoy (1992). Moreover, those relations have been shown to be very effective knowledge sources for WSD (McRoy 1992) and interpretation of noun sequences (Vanderwende 1994). For instance, *land*, *earth*, *mass*, *slope*, and *sand* are the genus terms that are categorically related to *bank*. On the other hand, words in the differentia such as river, lake, field, garden, bend, road, race-track, and harbour are situationally related to *bank* through the *Location* relation. Other keywords such as road, and race-track are related functionally to *bank* through the *PartOf* relation. For the most part, those relations exist conveniently among words under the same topic or across cross-referencing topics in LLOCE. For instance, most of the above mentioned words are listed under the same topic *Ld* (Geography) of the intended label *Ld099*, or its cross reference *Me* (Places). Therefore, these definitions can be disambiguated very effectively on the base of similarity between the defining keywords and the words lists in LLOCE.

2.1. Organizing information in LLOCE

In this work, the labels used for tagging dictionary definitions are taken from the LLOCE (McArthur 1992). Words in LLOCE are organized mainly according to subject matter. Nearly 2,500 sets of related words in LLOCE are organized according to 14 subjects and 129 topics (TOP). Cross references (REF) between sets, topics, and subjects are also given to show various inter-sense relations not captured within the same topic. The cross references in LLOCE are primarily between topics.

The sets under which the word is listed in LLOCE are considered as the initial candidates for labeling. For instance, the candidates for labeling senses of "bank" are the following 4 set labels:

- Je104* (banks, exchange, etc.),
- Je106* (banking and saving),
- Ld099* (river banks), and
- Nj295* (bending and leaning)

The set label *Je104* (as well as *Je106*) is listed under the topic *Je* (Banking, Wealth, and Investment), while

Ld099 and *Nj295* are listed under *Ld* (Geography) and *Nj* (Action and Position) respectively. For instance, there is a *REF* link (in Figure 1) from topic *Je* to topic *De* (Belonging and Owning, Getting and Giving). To facilitate estimation of similarity between a definition sentence and a topic, we use *TOPS* to denote the list of words under a LLOCE topic *S*, while *REFS* denotes the list of words under cross references of *S*. For instance, the label *Je104* (as well as *Je106*) is associated with a list of words from its topic (*TOPJe104*) and cross reference (*REFJe104 = TOPDe*):

TOPJe104 = TOPJe = {affluent, budget, cut down, deficit, economize, fortune, giro, income, keep, luxury, maintenance, needy, pay, windfall, amenity, ... }
REFJe104 = TOPDe = {bring back, contribution, doff, equip, facility, keep, yield, ... }.

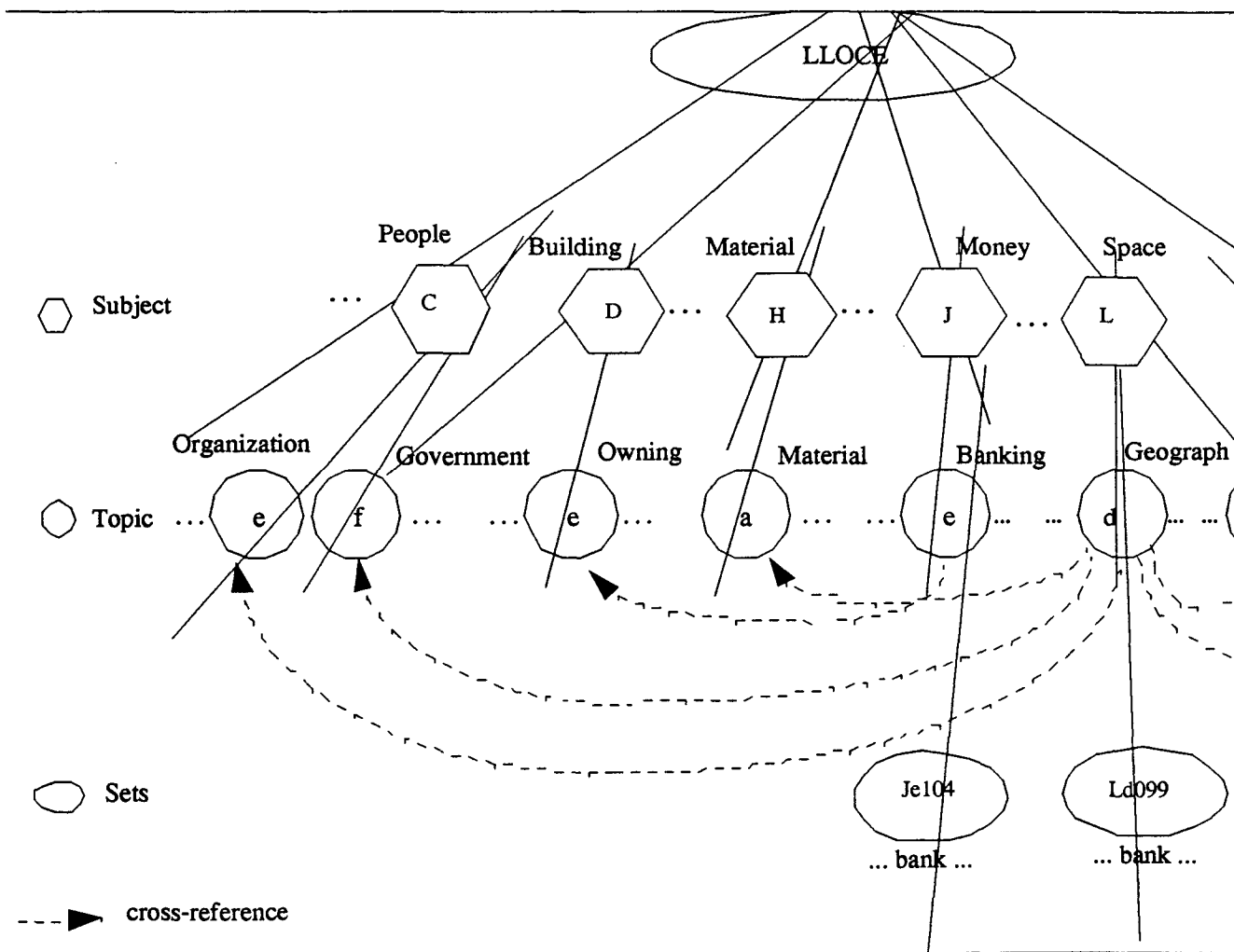


Figure 1. Subjects, topics, sets, and cross reference between topics in LLOCE.

3. The algorithm

The algorithm is divided into two stages. The preprocessing steps such as part-of-speech tagging, and removal of stop words are necessary for the algorithm to obtain good results. Various methods for POS tagging have been proposed in recent years. For simplicity, we adapted the method proposed by Church¹(1988) to tag the definition sentence. In the second stage, we select the label which is associated with word lists most similar to the definition as the result. We sum up the above descriptions and outline the procedure for labeling a dictionary sense.

Algorithm: Sense division for a head word h

Step 1: Given a head word h , read its definition, DEF_h , from LDOCE.

Step 2: For each definition D of DEF_h , tag each word in D with POS information..

Step 3: Remove all stop words in D to obtain a list of keyword-POS pair, $KEYD$.

Step 4: Lookup LLOCE for headword h to obtain a list of sets SET_h that contains h . For each S in SET_h , compile a set of words $TOPS$ that listed under the topic of S and $REFS$ the set of words listed under it cross references.

Step 5: Compute similarity $Sim(D, S)$ based on Dice Coefficient for all definitions $D \in DEF_h$ and labels $S \in SET_h$.

$$Sim(D, S) =$$

where $KEYD$ = the set of POS-keyword pairs in definition D ,

r = the overall relevancy of cross references to a topic,

w_k = 1/the degree of ambiguity of the keyword k ,

$In(a, B) = 1$ when $a \in B$,

$In(a, B) = 0$ when $a \notin B$.

Step 6: Assign to D the label S with the maximum value of $Sim(D, S)$ over a threshold.

Initially, the candidates are limited to the set labels indicated in LLOCE for the head word. If the algorithm finds all initial candidates dissimilar, a second run of the algorithm is executed with candidates expanded to all topics in LLOCE.

3.1 An illustrative example

We illustrate how the algorithm functions using the 5th definition of the word "interest." The preprocessing stage for definition of word "interest" includes part-of-speech (POS) tagging and stop word removal, thereby yielding the following result:

h	= "interest"
$SET_{interest}$	= {Fj228, Fb028, Je112, Ka006}
D	= "a share in a company business etc. "
$POSD$	= {a/det, share/n, in/prep, a/det, company/n, business/n, etc./adv}
$KEYD$	= {share/n, company/n, business/n}
$ KEYD $	= 3
$w_{share/n}$	= $1/ {De105, Hb037, Je114} = 1/3$
$w_{company/n}$	= $1/ {Cc042, Co292, Jh225} = 1/3$
$w_{business/n}$	= $1/ {Gh243, Jd138, Jh225} = 1/3$

1. In our case, tagging errors have very little negative impact, because words in LLOCE are organized primarily according to topic not part-of-speech.

$TOPFj228 = WFj = \{ \text{quite/adj, calm/adj, ... interest/n, excitement/n, shrill/n, ...} \}$
 $TOPFb028 = WFb = \{ \text{like/v, fancy/v, ... attraction/n, appeal/n, interest/n, ...} \}$
 $TOPJe112 = WJe = \{ \text{lend/v, loan/v, ... interest/n, investment/n, share/n, ...} \}$
 $TOPKa006 = WKa = \{ \text{entertain/v, amuse/v, ... game/n, hobby/n, interest/n, ...} \}$

$REFFj228 = WK = WKa \cup Wkb \dots \cup WKh$
 $= \{ \text{entertain/v, amuse/v, ... game/n, hobby/n, interest/n, ...} \}$
 $REFFb028 = Wcc = \{ \text{friend/n, acquaintance/n, ... companion/n, company/n, ...} \}$
 $REFJe112 = WDe = \{ \text{belong to/v, have/v, ... share/n} \}$
 $REFKa006 = WFj = \{ \text{quite/adj, calm/adj, ... interest/n, excitement/n, shrill/n, ...} \}$

$|TOPFj228 \cup REFFj228| = 1693$
 $|TOPFb028 \cup REFFb028| = 253$
 $|TOPJe112 \cup REFJe112| = 446$
 $|TOPKa006 \cup REFKa006| = 224$

$Sim(D, Fj228)^2 = 0$
 $Sim(D, Fb028) = 2 \cdot 0.33 \cdot (1 \cdot 1) / (253 + 3) = 0.66 / 256 = 0.00258$
 $Sim(D, Je112) = 2 \cdot 0.33 \cdot (1 + 1 \cdot 1) / (446 + 3) = 0.66 \cdot 2 / 449 = 0.00294$
 $Sim(D, Ka006) = 0$

The word lists associated with the label *Je112* is most similar to the key-words of the definition. Therefore, the algorithm produces *Je112* as the label for "a share in a company business etc."

3.2 Experiments and Evaluation

An experiment was carried out using a test set³ containing 12 polysemous words used in recent WSD experiments (Yarowsky 1992; Luk 1995). The 12-word test set used in the evaluation represents much more difficult cases than average. There are on the average 2.6 definitions in LDOCE for each words as opposed to the average 6.4 definitions per words in the test set. Table 1 displays a word by word performance of the algorithm. The results show that on the average the algorithm can assign labels to 87% of the senses with 94% precision.

4. Discussion

In this section, we thoroughly analyze the labeling performed by the algorithm and, in particular, look into several uses that are made possible by the labels' availability. In addition, those cases when the algorithm failed can also be analyzed. Analyses result not only illustrate the merits of these labels, but also imply possible improvement of the algorithm.

4.1. Broad coverage

About 50% of the labels are assigned during the second run of the algorithm from the extended candidate set. These labels represent gaps in the LLOCE. So, the algorithm can produce much broader coverage than the original LLOCE.

-
2. For simplicity, the parameter τ is set to 1.
 3. Only entries relevant to the test set in LLOCE are manually entered to the computer. We are currently trying to get a licence s.t. the full LLOCE entries in order to conduct a more complete test.

Table 1. Performance of the extended algorithm

headword	Alternate sets in LLOCE	No. of definitions in LDOCE	Labeling with expanded candidate set					
			Correct belling	La-	Incorrect Labelling	Unknown Labelling	Applicability	Precision
mole	n Ac061 n Mf159 n <u>Be</u> n <u>Ld</u>	3	1 - 1 1		0 - 0 0	0 - 0 0	100%	100%
sentence	n Cm256 n Gf235 v Gk210	3	1 1 1		0 0 0	0 0 0	100%	100%
slug	n Ag113 n Hh242 n <u>Gd</u> n <u>Hb</u> v <u>Nk335</u>	5	1 1 1 1 0		0 0 0 0 0	0 0 0 0 1	80%	100%
bass	adj Kb041 n Kb041 n <u>Kb</u> n <u>Ea</u>	7	1 3 1 1		0 0 0 0	0 1 0 0	86%	100%
bow	n Hh234 n Kb046 n Mf157 n Nj295 v Nj295 v <u>Mb</u> v <u>Kb</u> n <u>Dg</u> n <u>Na008</u>	11	1 1 1 1 2 1 1 0 0		0 0 0 1 0 0 0 1 0	0 0 0 0 0 0 0 0 1	91%	80%
cone	n Ai134 n Aj156 n Jb044 n <u>Ha</u>	3	- 1 1 1		- 0 0 0	- 0 0 0	100%	100%
duty	n Fc063 n Jf160 n <u>Jh</u>	2	- 1 1		- 0 0	- 0 0	100%	100%
galley	n Db038 n Mf157 n Mf153 n <u>Gd</u>	3	1 - 1 1		0 - 0 0	0 - 0 0	100%	100%

headword	Alternate sets in LLOCE	No. of definitions in LDOCE	Labeling with expanded candidate set				
			Correct labelling	La- labelling	Incorrect Labelling	Unknown Labelling	Applicabili- ty
interest	n Fb028	8	1	0	0	88%	100%
	n Fj228		-	-	-		
	n Ka006		1	0	0		
	n Je112		2	0	0		
	v Fb025		-	-	-		
	v Fj224		1	0	1		
	v Ka010		-	-	-		
	n <u>Nf</u>		2	0	0		
issue	n Aa020	10	-	-	-	80%	75%
	n Gd180		1	0	0		
	n Gf243		-	-	-		
	n Nf153		1	0	0		
	v Gd174		1	0	0		
	n <u>Ca</u>		1	0	0		
	v <u>De</u>		1	0	1		
	n <u>Na</u>		1	0	0		
	n <u>Ne</u>		0	2	0		
	n <u>Nb</u>		0	0	1		
star	n Kd082	11	1	0	0	73%	100%
	n La002		2	0	0		
	v Kd079		2	0	0		
	n <u>Hc</u>		1	0	0		
	n <u>Le</u>		1	0	0		
	n <u>Na</u>		1	0	0		
	n <u>Ia006</u>		0	0	1		
	n <u>Nb035</u>		0	0	1		
	n <u>Gd151</u>		0	0	1		
	taste		n Fl281	11	3		
v Fl280		-	-		-		
v,n <u>Ea</u>		4	0		0		
v,n <u>Fa</u>		2	0		0		
n <u>Gb</u>		1	0		0		
n <u>Fb020</u>		0	0		1		
total		77	63	4	10	87%	94%

Note: Extended labels are underscored.

4.2. Zero derivation

Dolan (1994) pointed out that it is helpful to identify zero-derived noun/verb pairs for such tasks as normalization of the semantics of expressions that are only superficially different. We have noticed that

zero derivatives are an important knowledge source for resolving PP-attachment ambiguity. A PP with an object involved in a noun/adjective zero-derivation has a strong tendency to attach itself to the preceding noun as a modifier. For instance, consider the following example that has an ambiguous PP-attachment problem:

We had a lot of interests in common.
(= *We had a lot of common interests.*)

4.3. Systematic inter-sense relations

Sanfilippo et al. (1995) contended that strong evidence would suggest that a large part of word sense ambiguity is not arbitrary but follows regular patterns. Moreover, gaps frequently arise in dictionaries and thesauri in specifying this kind of virtual polysemy. Virtual polysemy and recurring inter-sense relations are closely related to polymorphic senses that can support coercion in semantic typing under the theory of Generative Lexicon of Putstajovsky (1991).

Our experimental results indicate that the labels in LLOCE make it possible to acquire important inter-sense relations. Many of those relations are reflected in the cross reference information in LLOCE. For instance, LLOCE lists the following cross references for the topic of *Eb (Food)*:

Ac: Animals/Mammals
Ad: Birds
Af: Fish and other (water) creatures
Ah: Parts of animal
Ai: Kinds of parts of plants
Aj: Plant in general
Jg: Shopkeepers and shops selling food

most of which are systematic inter-sense relations similar to those described in above-mentioned work. We also observed that words involved in such inter-sense relations are frequently underspecified. For instance, "chicken" is listed under both topics *Eb* and topic *Ad*, while "duck" is listed under *Ad* but not *Eb*. By characterizing of some 200 cross references in LLOCE, most systematic inter-sense relations can be easily identified among the labeled senses. The labels attached to senses in the MRD, coupled with these inter-sense relations, can then support and realize automatic sense shifts advocated in Putstajovsky and Bouillon (1994). For instance, the sense of "duck" label with topic *Ad* can be coerced into an *Eb* sense when necessary, with the availability of the lexical rule stipulating a sense shift from *Ad* and *Eb*.

Krovetz (1992) observed that LDOCE indicates sense shifts via direct reference (links indicated by a capitalized word with a sense number) and deictic reference (implicit links to the previous sense created by *this*, *these*, *that*, *those*, *its*, *itself*, *such a*, *such an*). Sense shifts indicated through a deictic reference are also present in our 12-word test set. For instance, the first 2 senses of "issue" are

1. *the act of coming out.*
2. *an example of this.*

The definition of the 2nd senses indicates an *ActionNoun-CountNoun* sense shifts from *issue.n.1* to *issue.n.2* through a deictic reference of "this." Since those types of definitions pattern are not considered, the labeling algorithm fails in such cases. Further work must be undertaken to cope with direct and deictic references, so that such definitions can be appropriately labeled and information on sense shifts can be acquired.

4.4. Metonymy or Metaphor

Many definitions indicate metonymical or metaphorical associations between word senses. For instance, the 4th and 5th sense of "star" are

4. *a piece of metal in this shape for wearing as a mark of office, rank, honour, etc.*
5. *a heavenly body regarded as determining one 's fate.*

The 4th and 5th sense are metonymically associated with two "star" senses, *star.1.n.3* (a 5- or more pointed figure) and *start.1.n.2* (a heavenly body such as a PLANET), respectively. The algorithm often fails in such cases for two reasons. First, metonymies are not clearly separated and indicated in LLOCE. Second, the genus terms in metonymical senses are often indistinguishable from each other. Further action must be taken to identify the nature of such relations before this kind of ambiguity can be successfully resolved. The presence of phrases "as a ... of" or "regarded as" and drastic change in topic toward the second half of the definition may be cues for identifying metonymy and metaphor.

5. Other approaches

Sanfilippo and Poznanski (1992) proposed a so-called Dictionary Correlation Kit (DCK) in a dialog-based environment for correlating word senses across pairs of MRDs, LDOCE and LLOCE. Dolan (1994) described a heuristic approach to forming unlabeled clusters of closely related senses in a MRD. The clustering program relies on LDOCE domain code, grammar code, and 25 types of semantic relations extracted from definitions. Yarowsky (1992) described a WSD method and an implementation based on Roget's Thesaurus and the training material of the 10-million-word Grolier's Encyclopedia. The author suggested that the method can also apply to dictionary definitions. Krovetz (1993) described a simple algorithm based on overlap of defining words to identify related senses between morphological variants. The author reported that the success rate was over 80%. No results were reported for closely related senses within a part-of-speech.

In most of the above-mentioned works, experimental results are reported only for some senses of a couple of words. In this study, we have evaluated our method using all senses for 12 words that have been studied in WSD literature. This evaluation provides an overall picture for the expected success rate of the method, when applied to all word senses in the MRD. Directly comparing methods is often difficult. Nevertheless, it is evident that in comparison our algorithm is simpler, requires less preprocessing, and does not rely on information idiosyncratic to LDOCE. Thus, the algorithm described in this paper can readily apply to other MRDs besides LDOCE. Although our algorithm makes use of defining words with various semantic relations with the sense, explicit computation of those relations is not required.

6. Conclusions and Future Work

The method proposed in this work takes advantages of a number of linguistic phenomena: (1) Division of senses is primarily along the line of subject and topic. (2) Rather rigid schemes of text generation and predictable semantic relations are used to define senses in MRDs such as LDOCE. (3) The implicit links between instances of many of these relations are available in a thesaurus such as LLOCE.

This work also underscores the effectiveness of lexical rules for coarse WSD. Hand-constructed topic-based classes of words, coupled with lexical rules as common topic and cross references of topics, prove to be highly affective both in coverage and precision for WSD, admittedly for sense definitions, a somehow restricted type of text.

Merging senses via labeling has another implication as well. As discussed in Section 4, the senses sharing the same label (or cross-referencing labels) are frequently associated through various linguistic relations. Making those relations explicit will open the door to flexible treatment of lexicon, semantic typing, and semantic under-specification, all of which have received ever-increasing interest.

In a broader context, this paper promotes the progressive approach to knowledge acquisition for NLP as opposed to the "from-scratch" approach. We believe this to be a preferable means to approaching a sound and complete knowledge base.

Acknowledgment

The authors would like to thank the National Science Council of the ROC for financial support of this research under Contract No. NSC 85-2213-E-007-042.

References

- 1 . Brown, P., S.A. Pietra, V.J.D. Pietra, and R. Mercer (1991). "Word Sense Disambiguation using Statistical Methods," In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp 264-270.
- 2 . Church, Ken W. (1988). "A stochastic Parts Program and Noun Phrase Parser for Unrestricted Text." In *Proceedings of the 2nd Conference on Applied Natural Language Processing (ANLP-88)*, pp 136-143, Austin, Texas, USA.
- 3 . Dagan, Ido, Alon Itai, and Ulrike Schwall (1991). "Two Languages are More Informative than One," *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pp 130-137.
- 4 . Dagan, Ido, Alon Itai (1994), "Word Sense Disambiguation Using a Second Language Monolingual Corpus," *Computational Linguistics* 20(4), pp 563-596.
- 5 . Dolan, W.B. (1994). "Word Sense Disambiguation: Clustering Related Senses." In *Proceedings of the International Conference on Computational Linguistics*, pp 712-716.
- 6 . Kilgarriff, Adam (1993). "Dictionary Word Sense Distinctions: An Enquiry into Their Nature." *Computers and the Humanities*, 26, pp 365-387.
- 7 . Krovetz, R. and Croft (1992). "Lexical Ambiguity and Information Retrieval." *ACM Transaction on Information Systems*, pp 115-141.
- 8 . Krovetz, Robert (1992). "Sense-Linking in a Machine Readable Dictionary." In *Proceedings of the 30th Annual meeting of the Association for Computational Linguistics*, pp 330-332.
- 9 . Krovetz, Robert (1993). "Viewing Morphology as an Inference Process." In *Proceedings of the 16th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp 191-220.
- 10 . Lesk, Michael E. (1986). "Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from a ice-cream cone." In *Proceedings of the ACM SIGDOC Conference*, pp 24-26, Toronto, Ontario.
- 11 . Longman (1992). "Longman English-Chinese Dictionary of Contemporary English." *Longman Group (Far East) Ltd.*, Hong Kong.
- 12 . Luk, Alpha K (1995). "Statistical Sense Disambiguation with Relatively Small Corpora Using Dictionary Definitions." In *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp 181-188.
- 13 . McArthur, Tom (1992). "Longman Lexicon of Contemporary English." *Longman Group (Far East) Ltd.*, Hong Kong.
- 14 . McKeown, Katherine R. (1985). "Using Discourse and Focus Constraints to Generate Natural Language Text." *Cambridge University Press*, Cambridge, England.
- 15 . McRoy, S. (1992). "Using Multiple Knowledge Sources for Word Sense Discrimination." *Computational Linguistics* 18(1), pp 1-30.
- 16 . Putstejovsky, James (1991). "The Generative Lexicon." *Computational Linguistics* (17)4, pp 409-441.
- 17 . Putstejovsky, James and Pierrette Bouillon (1994). "On the Proper Role of Coercion in Semantic Typing." In *Proceedings of the International Conference on Computational Linguistics*, pp 706-711.
- 18 . Sanfilippo, A., and V. Poznanski. (1992). "The Acquisition of Lexical Knowledge from Combined

Machine-Readable Dictionary Sources." In *Proceedings of the 3rd Conference on Applied Natural Language Processing (ANLP-92)*, pp 80-87, Trento, Italy.

19. Vanderwende (1994). "Interpretation of Noun Sequence." In *Proceedings of the International Conference on Computational Linguistics*, pp 454-460.

20. Yarowsky, David (1992). "Word-Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora." In *Proceedings of the International Conference on Computational Linguistics*, pp 454-460.

Appendix Semantic Table for set labels of 12 Polysemous Words

Word	Alternate sets in LLOCE	Semantic Labelling
mole	n Ac061 n Mf159 n Be n Ld	small animals harbours and yards skin, complexion and hair geography
sentence	n Cm256 n Gf235 v Ck210	punishments and deterrents phrases and sentences punishing and fining
slug	n Ag113 n Hh242 n Gd n Hb n Nk335	worms and similar creatures shot and bullets communicating object generally striking
bow	n Hh234 n Kb046 n Mf157 n Nj295 v Nj295 v Mb v Kb n Dg n Na008	bows and arrows string instruments parts of ships bending and leaning bending and leaning putting and taking music and related activities clothes and personal belongings typifying and embodying
cone	n Ai134 n Aj156 n Jb044 n Ha	fruit and seed kinds of coniferous trees geometrical shapes substances, materials, objects and equipment
duty	n Fc063 n Jf160 n Jh	rules of behaviour taxes business, work, and employment
galley	n Db038 n Mf157 n Mf153 n Gd	the kitchen and similar rooms parts of ships larger kinds of sailing boats communication

Word	Alternate sets in LLOCE	Semantic Labelling
interest	n Fb028 n Fj228 n Ka006 n Je112 v Fb025 v Fj224 v Ka010 n Nf	attracting communicating, mainly by reading and writing, printing and publishing games and hobbies interest on money attracting and interesting interesting and exciting interesting and thrilling causing
issue	n Aa020 n Gd180 n Gf243 n Nf153 v Gd174 n Ca v De n Na n Nb n Ne	young creatures publications and editions subject and topic results and effects publishing people belonging and owning, getting and giving being, becoming and happening chance doing things
star	n Kd082 n La002 v Kd079 n Hc n Le n Na n Ia006 n Nb035 n Gd151	actors planets, suns, and stars playing and rehearsing specific substances and materials time generally being, becoming and happening shapes and models fortune punctuation
taste	n F1281 v F1280 v,n Ea v,n Fa n Gb n Fb020	tasting things tasting things food generally feeling and behavior generally knowing and learning liking and loving