Adaptive Word Sense Disambiguation Using Lexical Knowledge in a Machine-readable Dictionary

Jen Nan Chen[*]

Abstract

This paper describes a general framework for adaptive conceptual word sense disambiguation. The proposed system begins with knowledge acquisition from machine-readable dictionaries. Central to the approach is the adaptive step that enriches the initial knowledge base with knowledge gleaned from the partial disambiguated text. Once the knowledge base is adjusted to suit the text at hand, it is applied to the text again to finalize the disambiguation decision. Definitions and example sentences from the Longman Dictionary of Contemporary English are employed as training materials for word sense disambiguation, while passages from the Brown corpus and Wall Street Journal (WSJ) articles are used for testing. An experiment showed that adaptation did significantly improve the success rate. For thirteen highly ambiguous words, the proposed method disambiguated with an average precision rate of 70.5% for the Brown corpus and 77.3% for the WSJ articles.

Keywords: word sense disambiguation, machine-readable dictionary, semantics.

1. Introduction

Word sense disambiguation is a long-standing problem in natural language understanding. It seems to be very difficult to statistically acquire enough word-based knowledge about a language to build a robust system capable of automatically disambiguating senses in unrestricted text. For such a system to be effective, a large number of balanced materials must be assembled in order to cover many idiosyncratic aspects of the language. There exist three issues in a lexicalized statistical word sense disambiguation (WSD) model: data sparseness, the lack of abstraction, and static learning. First, a word-based model has a plethora of parameters that are difficult to estimate reliably even with a very large corpus. Under-trained models lead to low precision. Second, word-based models lack a degree of abstraction

---

[*] Department of Information Management, Ming Chuan University, Shih-lin, Taiwan, R.O.C

that is crucial for a broad coverage system.   Third, a static WSD model is unlikely to be robust and portable, since it is very difficult to build a single model relevant to a wide variety of unrestricted texts. Several WSD systems have been developed that apply word-based models to a specific or genre domain to disambiguate senses appearing in generally easy context that has a large number of typically salient words.   In the case of unrestricted text, however, the context tends to be very diverse and difficult to capture with a lexicalized model; therefore, a corpus-trained system is unlikely to transfer well to a new domain.

Generality and adaptability are, therefore, keys to a robust and portable WSD system.   A concept-based model for WSD requires fewer parameters and has an element of generality built in. Conceptual classes make it possible to generalize from word-specific context in order to disambiguate word senses appearing in an unfamiliar context in terms of word recurrences.   An adaptive system, armed with an initial lexical and conceptual knowledge base extracted from machine-readable dictionaries (MRD), has two strong advantages over static lexicalized models trained on a corpus.   First, the initial knowledge is rich and unbiased enough for a substantial portion of text to be disambiguated correctly. Second, based on the result of initial disambiguation, an adaptation step can then be performed to make the knowledge base more relevant to the task at hand, thus resulting in broader and more precise WSD.

In this paper we explore in some depth the question of whether conceptual knowledge in the MRD is effective enough to provide a general solution for disambiguating contexts of unrestricted texts, such as the Brown and Wall Street Journal (WSJ) corpora.   Major emphasis has previously been placed on self-adaptation [Chen and Chang 1998a].   This approach is based on the hypothesis that a substantial part of a given text is *easy* or prototypical and, therefore, susceptible to interpretation based on general knowledge derived from the MRD.   By adapting the contextual representation of word senses to those in the easy context, we hope to be better equipped to interpret the other part, which is usually considered a *hard* context.   Adaptation results in gaps in the general knowledge being filled in or domain specific information being added to the initial knowledge base.   Either way, adaptation makes the knowledge base more relevant to the text and, therefore, more effective for WSD in a hard context.   We will give experimental results showing the effectiveness of this adaptive WSD approach based on initial knowledge base acquired from the MRD.   Although our adaptive approach requires virtually no domain-specific training, it nevertheless achieves high precision rates for WSD of unrestricted text rivaling those of static methods that demand very lengthy training using a very large corpus.

Figure 1 lays out the general framework for the adaptive conceptual WSD approach which this research employed.   The learning process described here begins with a step involving knowledge acquisition from MRDs.   With this acquired knowledge, the input text is read and a trial disambiguation step is carried out.   An adaptation step follows which combines the initial knowledge base with knowledge gleaned from the partially disambiguated text.   Once the knowledge base is adjusted to suit the text at hand, it is then applied to the text again to finalize the disambiguation result.   For instance, the

initial contextual representation (CR) extracted from the Longman Dictionary of Contempory English [Proctor 1978, LDOCE] for the *bank*-GEO sense contains both lexical and conceptual information: {*land*, *river*, *lake*, ⋯} ∪ {GEO, MOTION, ⋯}.   The initial CR is informative enough to disambiguate a passage containing "*a deer near the river bank*" in the input text.   The trial disambiguation step produces sense tagging of *deer*/ANIMAL and *bank*/GEO, but certain instances of *bank* are left untagged due to the lack of WSD knowledge.   We observe that the *bank*-GEO sense in the context of *vole* is unresolved since there is no link between ANIMAL and GEOGRAPHY.   Subsequently, the adaptation step adds *deer* and ANIMAL to the contextual representation for *bank*-GEO.   The adapted CR is now enriched with information capable of disambiguating the instance of *bank* in the context of *vole* to produce the final disambiguation result.

The rest of this paper is organized as follows.   First of all, we will peresent how easy contexts are interpreted and ambiguous words are labeled in the initial disambiguation step using general knowledge derived from MRD.   Next, we describe the adaptation step that uses the sense labels assigned to polysemous words.   After that, we will describe the strategy of using the adapted knowledge base and defaults.   Next, we will give a detailed account of experiments conducted to assess the effectiveness of the adaptive approach, including the experiment setup, results and evaluation.   Following that, we will review the recent WSD literature from the perspective of various types of contextual knowledge and different representation schemes.   Finally, we will draw conclusions.
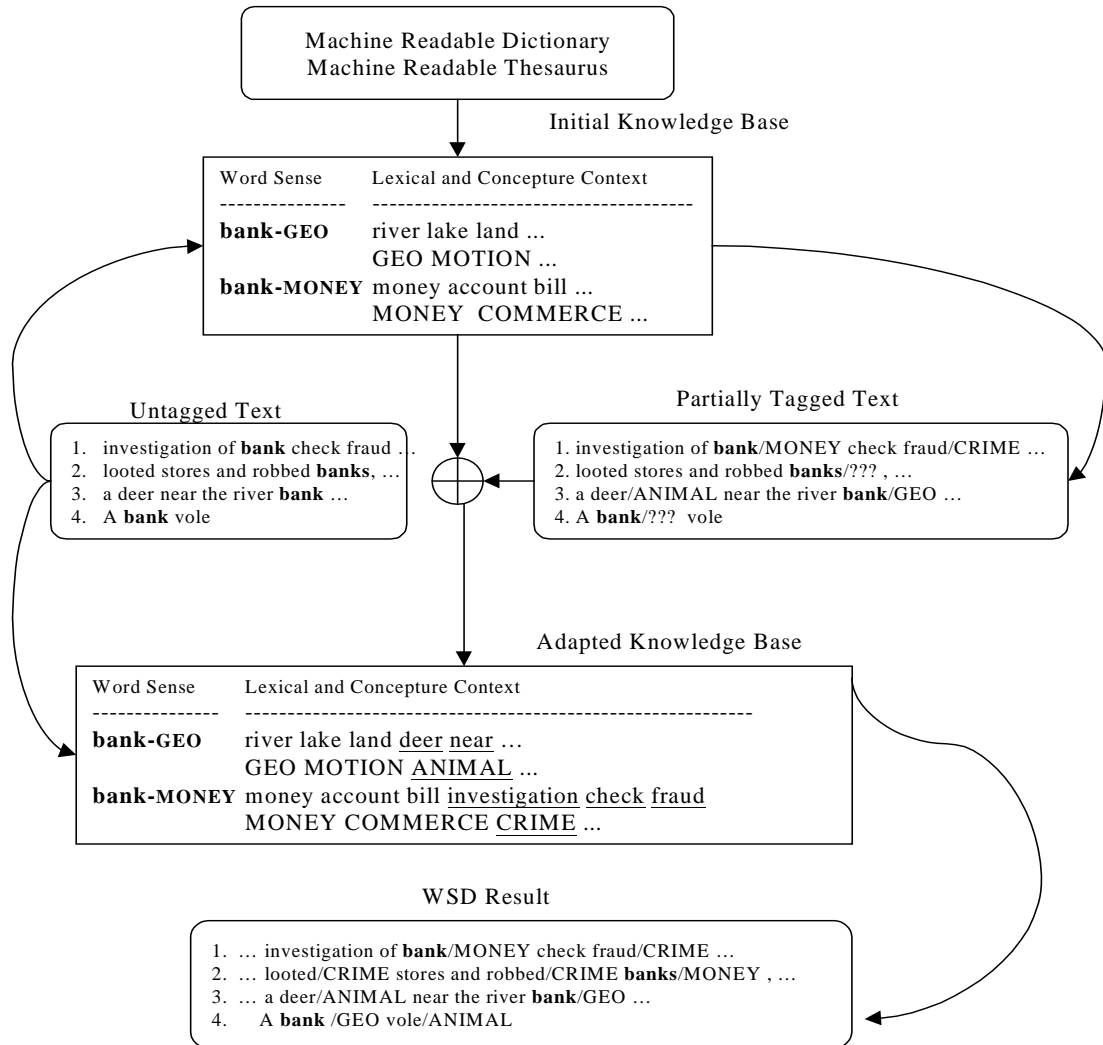
Figure 1 *General framework for adaptive WSD using MRD.*

2. Acquisition of Disambiguation Knowledge using MRD

In this section, we will describe how the conceptual characterization technique is applied to MRD definitions and give examples of acquiring WSD knowledge. First, we will show word level definitions based on a lexical CR and then a conceptual CR. Next, we will show the advantage of including information gained from an example sentence. Finally, we will combine these techniques to perform adaptative WSD computation.

2.1 Contextual Representation from Sense Definition

2.1.1 Lexicalized Contextual Representation

A word-level contextual representation from MRD definitions can be derived almost effortlessly. Let $CR(W, S)$ denote the contextual representation of the sense S of headword W. Intuitively, it is composed of the content words in the definition with a specific sense. Thus, $CR(W, S)$ can be represented symbolically as { $x \mid x \in DEF^1_{W, S}$ and x is not a function word }. To illustrate, the CRs for the nine nominal senses of *bank* in LDOCE listed below are shown in Table 1:

bank.1.n.1	land along the side of a river, lake, etc.;
bank.1.n.2	earth which is heaped up in a field or garden, often making a border or division;
bank.1.n.3	a mass of snow, clouds, mud, etc.;
bank.1.n.4	a slope made at bends in a road or race-track, so that they are safer for cars to go round;
bank.1.n.5	a high underwater of bank in a river, harbour, etc.;
bank.3.n.1	a row, esp. of OAR in an ancient boat or KEY on a TYPEWRITER;
bank.4.n.1	a place in which money is kept and paid out on demand, and where related activities go on;
bank.4.n.2	a place where something is held ready for use, esp. ORGANIC products of human origin for medical use;
bank.4.n.3	a person who keeps a supply of money or pieces for payment or use in a game of chance.

Table 1. *Lexical contextual representations for* bank *senses.*

Sense ID	Sense Label S	Lexical Context Representation $LCR(D_{bank, s})$
bank.4.n.1	MONEY	{place, money, keep, pay, demand, activity}
bank.1.n.1	RIVER	{land, lake, river}
bank.1.n.5	SANDBANK	{underwater, sand, harbour}
bank.1.n.2	EARTH	{earth, heap, field, garden, boarder, division}
bank.1.n.3	PILE	{mass, snow, cloud, mud}
bank.1.n.4	ROAD	{car, aircraft, move, side, turn}
bank.3.n.1	ROW	{row, oar, boat, key, typewriter}
bank.4.n.2	MEDICINE	{place, hold, use, organic, product, human, origin, medical}
bank.4.n.3	GAMBLE	{person, keep, supply, money, payment, game, chance}

2.1.2 Conceptualized Contextual Representation

The word-based CR from MRD definitions is highly precise and effective but not broad enough to work alone effectively. Word-based sense representation is hampered by the difficulty of providing estimates for a very large parameter space leading to limited coverage in WSD. Certainly, there are many

---

[1] By $DEF_{S, W}$, we mean the definition of sense S of headword W in MRD.

situations that call for a conceptual generalization of a word-based representation of word sense from an example sentence.   For instance, the RIVER sense of *bank* in Example (1c) can be correctly interpreted by an MRD-based CR, but only when the contextual word *river* in the CR is generalized to all words related to RIVER, including the word *stream*:

(1) a. a ribbon of mist along the river <u>bank</u>;

     b. a small excavation in the river <u>bank</u>;

     c. the left <u>bank</u> of the stream.

There are many possible approaches to making such a generalization and deriving a conceptualized CR (CCR) of word sense.   Chen and Chang [1998b] described one such approach based on thesaurus topics. The CCR for each MRD sense can be viewed as relating to words listed under some Longman Lexicon of Contemporary English [McArthur 1992, LLOCE] topics.   By linking MRD senses to thesaurus senses and by classifying senses according to linked senses, we can derive the CCR for a sense definition.   Table 2 shows the topical CCR for the senses of *bank* in LDOCE.   Each sense in MRD is given a list of weighted LLOCE topics.   The weights in the CCR are normalized to a sum of unity for the obvious reason.   Table 3 shows the lists of words listed in LLOCE under the topics relevant to *bank* senses.

Table 2. Conceptual context representations for a definition D *of* bank *senses.*

Sense Label S	Topics[2] (with weights) on $CCR(D_{bank,s})$
MONEY	Je(0.45), Jf(0.33), Jd(0.22)
RIVER	Ld(0.45), Mf(0.26), Me(0.14), Hc(0.07), Af(0.05), Ad(0.04)
EARTH	La(0.36), Ld(0.24), Eg(0.20), Me(0.12), Ie(0.08)
PILE	Lc(0.59), Db(0.13), Hc(0.09), La(0.09), Md(0.09)
ROAD	Md(0.45), Me(0.38), Ld(0.17)
ROW	Md(0.49), Gd(0.18), Mc(0.16), Kb(0.12), Me(0.06)
MEDICINE	Bd(0.70), Bj(0.30)
GAMBLE	Ke(0.35), Kh(0.28), Kf(0.23), Cn(0.14)

We sum up the above description and outline the procedure as Algorithm 1 for creating a CCR for a word W of sense S with definition D.

Algorithm 1: Creating a conceptualized contextual representation $CCR(D_{w,s})$ for a word W of sense S with definition D.

Step 1: Run the *TopSense* algorithm described by Chen and Chang [1998b] to map D to *SC(D)*, a set of semantic categories in a thesaurus.

---

[2] See Appendix A for more details.

Step 2: Create a conceptualized contextual representation $CCR(D_{w,s})$ for sense S with definition D:

$$CCR(D_{w,s}) = \sum_{T \in SC(D)} WORD(T)^3,$$

where *WORD(T)* is a set of related words in semantic category *T*.

Table 3. *Definition-based conceptual context representation and related word lists for* bank *senses*

Sense Division S	Topics	Word List on $CCR(D_{bank,s})$
MONEY	Je(0.45)	{money, pay, cash, capital, account, charge, ...
	Jf(0.33)	pay, bond, bill, charge, ...
	Jd(0.22)	money, cash, fund, check, ... }
RIVER	Ld(0.45)	{lake, land, river, shore, stream, beach, ...
	Mf(0.26)	boat, ship, craft, port, ...
	Me(0.14)	place, edge, road, border, ...
	Hc(0.07)	rock, stone, clay, soil, ...
	Af(0.05)	fish, crab, coral, shell, fur, ...
	Ad(0.04)	chicken, duck, goose, seabird, ...}
EARTH	La(0.36)	{universe, space, planet, constellation, ...
	Ld(0.24)	lake, land, river, shore, stream, beach, ...
	Eg(0.20)	farming, field, crop, stock, productive, ...
	Me(0.12)	place, edge, road, border, ...
	Ie(0.08)	playgroup, school, college, classroom, ...}
PILE	Lc(0.59)	{weather, climate, sky, cloud, fog, steam, ...
	Db(0.13)	roof, ceiling, wall, door, ground, ...
	Hc(0.09)	rock, stone, clay, soil, ...
	La(0.09)	universe, space, planet, constellation, ...
	Md(0.09)	transport, vehicle, car, motorcar, transit, ...}
ROAD	Md(0.45)	{transport, vehicle, car, motorcar, transit, ...
	Me(0.38)	place, edge, road, border, ...
	Ld(0.17)	lake, land, river, shore, stream, beach, ...}
ROW	Md(0.49)	{transport, vehicle, car, motorcar, transit, ...
	Gd(0.18)	printing, sign, letter, code, …
	Mc(0.16)	sail, caravan, itinerary, …
	Kb(0.12)	song, melody, dance, …
	Me(0.06)	road, street, …}
MEDICINE	Bd(0.70)	{blood, trunk, breast, back, buttock, waist, ...
	Bj(0.30)	patient, examine, diagnose, soothe, ...}
GAMBLE	Ke(0.35)	{athletics, run, jump, ride, game, round, ...
	Kh(0.28)	ball game, shoot, golf, pitch, football, ...

[3] *WORD*(*T*) is a bag of words rather than a set. By summation of bags, we mean collecting all the word instances in the bags and keeping track of counts.

Sense Division S	Topics	Word List on $CCR(D_{bank, s})$
MONEY	Je(0.45)	{money, pay, cash, capital, account, charge, ...
	Kf(0.23)	cards, pack, suit, heart, club, poker, dice, ...
	Cn(0.14)	war, warfare, conflict, fight, ...}

An Illustrative Example

In the following, we demonstrate how Algorithm 1 works. Given the sense definition of bank.4.n.1 shown in Section 2.1.1, the *CCR*(bank.4.n.1) can be acquired by means of Algorithm 1.

Step 1: After running the *TopSense* algorithm, we have *SC*(bank.4.n.1) = {**Je**, **Jf**, **Jd**}.

Step 2: Next, we expand each of three topics in *SC*(bank.4.n.1) to a cluster of words. Thus, we have

> *WORD* (**Je**) ={money, pay, cash, capital, account, charge, ...},
>
> *WORD* (**Jf**) ={pay, bond, bill, charge, ...} and
>
> *WORD* (**Jd**) ={money, cash, fund, check, ...}.
>
> Finally, the *CCR* (**bank.4.n.1**) =*WORD* (**Je**) + *WORD* (**Jf**) + *WORD* (**Jd**)
> > ={money, pay, cash, capital, ...,
> > pay, bond, bill, charge, ...,
> > money, cash, fund, check, ... }.

2.2 Contextual Representation from an Example Sentence

Dictionary examples are intended to show typical use of words in context. Therefore, MRD examples provide rich information supplementary to definitions. In this section, we will describe a method for tagging bilingual sentences with sense labels based on dictionary definitions and translations in a bilingual MRD.

However, the sense for each word in an example is not explicitly marked except for the word being defined. That limits the potential for using dictionary examples as knowledge sources for WSD. Gale, Church and Yarowsky [1992b] first pointed out that the strong constraint of one-sense-per-translation can be exploited to tag a bilingual corpus for training a statistical WSD model. Building on their idea, we describe a new method for tagging bilingual sentences, in the MRD or elsewhere, for automatic acquisition of the CR of senses.

2.2.1 Sense Tagging Based on Conceptual Context Representation and Translations

Now we are ready to propose a heuristic algorithm for tagging bilingual sentences with sense labels. First, the translation morphemes of an MRD definition are added to the CR so that not only the English context, but also the translation (in Chinese for the particular implementation of LDOCE/E-C we will be describing) is considered. For instance, the representation of MONEY-bank contains not only FINANCE words such as *money*, *pay*, *cash*, *capital*, *account*, *charge*, etc., but also the morphemes "銀"

and "行" in the translation of the definition. (See Table 4 for some examples of bilingual context representations for the *bank* senses in LDOCE/E-C.) Subsequently, each CR for a polysemous word is compared with the bilingual sentences. The polysemous word is tagged in favor of the relevant CR that has the most overlap with the bilingual sentences. For instance, consider the case of tagging the instance of *bank* in Example (2) extracted from LDOCE/E-C:

(2) a. the interest in my bank account accrued over the years;
    b. 我銀行帳戶的利息逐年有所增加。

Under the assumption of the one-sense-per-translation constraint, the morphemes "銀" and "行" in the translation are sufficient evidence for tagging the instance of *bank* as MONEY-bank. Even if such telling evidence is not present, there nevertheless is a great chance that the sentence contains enough words related to a relevant topic for correct sense tagging to happen. For instance, the FINANCE words, such as *interest* and *account*, in Example (2) lead to the correct sense label MONEY-bank for this instance of *bank*, even when it is not translated as "銀行." The contextual representation derived from the MRD definition also acts as a safety net when the one-translation-per-sense constraint does not hold. For instance, based on the one-translation-per-sense constraint, the instance of *star* in Example (3) can not be labeled as ENTERTAINMENT-star because both the ENTERTAINMENT and HEAVENLY-BODY senses of *star* are translated as "星":

(3) a. she is a star with the theatre company;
    b. 她是劇團的紅星。

In such an event, the ENTERTAINMENT words, such as *theatre* and *company*, nonetheless result in the correct sense label: ENTERTAINMENT-star.

A bilingual example in the MRD, or text in a bilingual corpus, can be tagged in the way described above, word by word and sentence by sentence. Unambiguous words with only one sense label are tagged as such. Tagging is done only for content words within the scope of this work. Function words can be treated similarly [Chang, Hsu and Chen 1996]. Sentences in English tagged as training materials can facilitate acquisition of WSD knowledge. The method for tagging a bilingual training corpus is summarized as Algorithm 2. Table 5 shows the result of applying Algorithm 2 to some LDOCE/E-C examples.

Table 4. *Bilingual contextual representations for* bank *senses based on conceptual context representations from definition and dictionary translation.*

Sense Division S	Context on $CCR(D_{bank, s})$
MONEY	{銀，行， money, pay, cash, capital, account, charge ..., pay, bond, bill, charge ..., money, cash, fund, check, ... }
RIVER	{岸，堤，沙，洲, lake, land, river, shore, stream, beach, ..., boat, ship, craft, port, ..., place, edge, road, border, ..., rock, stone, clay, soil, ..., fish, crab, coral, shell, fur, ..., chicken, duct, goose, seabird, ... }
EARTH	{田，埂， universe, space, planet, constellation, ..., lake, land, river, shore, stream, beach, ..., farming, field, crop, stock, productive, ..., place, edge, road, border, ..., playgroup, school, college, classroom, ... }
PILE	{一，塊，一，團， weather, climate, sky, cloud, fog, steam, ..., roof, chimney, ceiling, wall, door, ground, ..., rock, stone, clay, soil, ..., universe, space, planet, constellation, …, transport, vehicle, car, motorcar, transit, ... }
ROAD	{邊，坡， transport, vehicle, car, motorcar, transit, ..., place, edge, road, border, ..., lake, land, river, shore, stream, beach, ... }
ROW	{一，排， transport, vehicle, car, motorcar, transit, … , printing, sign, letter, code, …, sail, caravan, itinerary,…, song, melody, dance, …, road, street, ……... }
MEDICINE	{血，庫， blood, trunk, breast, back, buttock, waist, ..., patient, examine, diagnose, soothe, ... }
GAMBLE	{莊，家， athletics, run, jump, ride, game, round, ..., ball game, shoot, golf, pitch, football, ..., cards, pack, suit, heart, club, poker, dice, ... war, warfare, conflict, fight, battleground, ... }

Table 5. *Results of sense tagging.*

Example	The interest in my bank account accrued over the years.
Translation	我銀行帳戶的利息逐年有所增加。
Tagged Keywords	interest/**Je**, bank/**Je**, account/**Je**, accrue/**Nd**

Gloss for Topics	Je	Banking
	Nd	Size

Algorithm 2: Labeling bilingual training corpus

Step 1: Form contextual representation *CR(W, S)* of sense *S* of word *W* with definition *D* and translation *T* as follows:

$$CR(W, S) = LCR(D_{w, s}) + CCR(D_{w, s}) + LCR(T_{w, s}).$$

Step 2: For each word W in an example sentence E, compute the similarity between its context and translation, C_E, and each of the contextual representations $CR(W,S)$ based on the Dice Coefficient:

$$\text{Sim}(C_E, CR(W, S)) \quad = \quad \sum_{c \in C_E} \frac{2 \times \text{In}(C, CR(W, S))}{|C_E| + |CR(W, S)|},$$

where $\text{In}(a, B)$ = the weight of a in B, if $a \in B$ and

0, otherwise.

Step 3: Label W in E with S^* such that $\text{Sim}(C_E, CR(W,S^*))$ is maximized;

$$\text{Sim}(C_E, CR(W,S^*)) = \underset{L}{\text{Max}}\ \text{Sim}(C_E, CR(W,L)) \text{ and is greater than a certain threshold.}$$

2.2.2 Acquiring Contextual Representations for Example Sentences

Lexicalized and conceptualized CR can be constructed from tagged MRD examples in a fashion similar to that described in Section 2.1 for MRD definitions. Given an ambiguous word W labeled with sense S in a set of example sentences $E_{w,s}$, every content word appearing in E is gathered to form $LCR(E_{w,s})$, shown as follows:

$$LCR(E_{w,s}) = \{\, x \mid x \in E_{w,s} \text{ and } x \text{ is not a function word} \,\}.$$

Table 6 shows some of the contextual words in the LDOCE examples that appear in the context of each of eight *bank* senses. Notice that the entry for MONEY-bank contains many strong collocates, such as (*rob*, *bank*), (*bank*, *account*), etc. These collocates are potentially very helpful for WSD. Although some of the contextual words merely repeat information in the definition-based representation, $LCR(D_{w,s}) + CCR(D_{w,s})$, many do provide new information. For instance, fifteen instances of *river* reaffirm the defining word *river* as an important collocate for RIVER-bank, while contextual words such as *north*, *east*, *deer*, and *vole* provide additional, richer context.

Table 6. *Lexicalized contextual representations for* bank *from the set of LDOCE examples* E.

Sense Label S	Context (with frequency) in $LCR(E_{\text{bank},s})$
MONEY	rob(23), account(15), money(8), criminal(6), interest(5), keep(5), paper(4), police(4), robber(4), thief(4), cheque(3), ...
RIVER	river(15), city(2), north(2), stream(2), east(1), air(1), deer(1), south(1), sea(1), vole(1), ...
EARTH	build(2), earth(2), flood(1), rise(1), water(1), ...
PILE	cloud(2), dark(2), heavy(1), storm(1), ...
ROAD	moss(2), wood(2), rest(1), sit(1), ...
ROW	-
MEDICINE	-
GAMBLE	-

In the previous section, we showed that these contextual words are neither frequent nor necessarily likely to recur. However, when viewed as representing a typical topic or concept, they certainly are recurring. For instance, although there is only one instance of *north bank* in LDOCE examples, there are quite a few *south bank*, and *right bank* instances, all of which signal a recurring context of the DIRECTION concept. Therefore, it is a good idea to derive a conceptualized contextual representation from the set of examples *E* relevant to a sense label *S*. For instance, representing the co-occurring concept of the DIRECTION with RIVER- bank, $CCR(E_{bank, river})$ would contain such words as *east, west, south, north, left, and right*, etc.:

$$CCR(E_{bank, river}) = \{ \textit{ east, west, south, north, left, right, } \dots \}.$$

For this purpose, we again turn to the information retrieval (IR) technique. Since the LDOCE in general strictly uses words in the controlled vocabulary for both definitions and examples, the same method described by Chen and Chang [1998b] for forming conceptual characterization of MRD definitions also works for MRD examples. Table 7 shows a list of topical words that characterize the context of each of the eight *bank* senses based on sense tagged LDOCE examples. The results obtained using an IR-based method seem to characterize the context in a general way that can be very useful for WSD.

Table 7. *Conceptualized contextual representation for* bank *from the set of LDOCE examples E.*

Sense Division *S*	Related Topics	Context on $CCR(E_{bank, s})$
MONEY	Je(0.45), Jf(0.33), Jd(0.22)	{officer, cop, detective, guard, protect, gangster, hoodlum, larceny, hijacking, burglar, steal, fraud, swindle, …}
RIVER	Ld(0.45), Mf(0.26), Me(0.14), Hc(0.07), Af(0.05), Ad(0.04)	{east, west, north, south, up, down, erode, elk, moose, rat, mouse, rabbit, hare, … }
EARTH	La(0.36), Ld(0.24), Eg(0.20), Me(0.12), Ie(0.08)	{tide, ebb, current, spate, …}
PILE	Lc(0.59), Db(0.13), Hc(0.09), La(0.09), Md(0.09)	{fog, steam, haze, dew, mist, …}
ROAD	Md(0.45), Me(0.38), Ld(0.17)	{forest, jungle, hole, crack, …}
ROW	Md(0.49), Gd(0.18), Mc(0.16), Kb(0.12), Me(0.06)	-
MEDICINE	Bd(0.70), Bj(0.30)	-
GAMBLE	Ke(0.35), Kh(0.28), Kf(0.23), Cn(0.14)	-

2.3 Combining Definition-based and Example-based CR

Definition-based and example-based CR as described in Sections 2.1 and 2.2 can be put together to form a combined CR for acquiring word sense. For simplicity, we merge the two to produce the final MRD-based CR. For a polysemous word *W* and a relevant word sense *S*, with the definition *D* of sense

S and the set of examples *E* containing an instance of *S*, the contextual representation *CR*(*W, S*) can be represented as follows:

$$WORD(W, S) = LCR(D_{w,s}) + CCR(D_{w,s}) + LCR(E_{w,s}) + CCR(E_{w,s}),$$

where $LCR(D_{w,s})$ is the lexical contextual representation derived from definition *D*,

$CCR(D_{w,s})$ is the conceptual contextual representation derived from definition *D*,

$LCR(E_{w,s})$ is the lexical contextual representation derived from the set of examples *E*,

and $CCR(E_{w,s})$ is the conceptual contextual representation derived from the set of examples *E*.

To take into account the significance of each contextual word in *CR*(*W, S*), the IR technique for weighting index terms for relevancy can be applied here to good effect. Using the IR analogy, the collective context of each word sense is viewed as a document, and the relevance of a contextual word *t* to a sense *S* of word *W* depends on its term frequency *tf* and inverse document frequency *idf*. The term frequency *tf* is the number of instances of *t* in *WORD*(*W, S*), and *idf* is the percentage of *CR*s in which an instance of *t* appears. The relevancy of a contextual word is estimated using the commonly used scheme: *tf* $\times idf$. Experiments show that the simple scheme tends to give a high weight to strong collocations, such as (*rob*, MONEY-bank) and (*river*, RIVER-bank), thus leading to a representation that is potentially very effective for WSD.

We sum up the above descriptions and outline the procedure as Algorithm 3. The algorithm combines definition-based and example-based CR into an integrated contextual representation *CR*(*W, S*) for the sense *S* of the polysemous word *W*.

Algorithm 3: Combining definition-based CR

Step 1: Given a polysemous word *W*, one of its senses *S* and a collection of bilingual examples *C*, run Algorithms 1 and 2 to obtain $LCR(D_{w,s})$, $CCR(D_{w,s})$, $LCR(E_{w,s})$ and $CCR(E_{w,s})$, where *E* is a set of examples that each contain an instance of *S*.

Step 2: Merge the following word list for W and *S*:
$$WORD(W, S) = LCR(D_{w,s}) + CCR(D_{w,s}) + LCR(E_{w,s}) + CCR(E_{w,s}).$$

Step 3: For each *WORD*(*W, S*), compute a list of distinct words *X* with weight $W_{X,S}$ as follows:
$CR(W, S) = \{ X (W_{X,S}) \mid X$ is a distinct word in *WORD*(*W, S*)},
where $tf_{X,S}$ = the frequency of *X* in *WORD*(*W, S*),
idf_X = 1/the percentage of senses *S* such that $X \in WORD(W, S)$,
$W_{X,S}$ = $tf_{X,S} \times idf_X$.

Step 4: The weights $W_{X,S}$ in *CR*(*W, S*) for each word sense *S* are normalized to a sum of 100.

An Illustrative Example

In the following, we will demonstrate how Algorithm 3 works. Given a MONEY-bank sense, the integrated *CR*(*bank*, MONEY) can be acquired by doing the following (where the numbers in parentheses following collocates denote the frequency):

Step 1: After running Algorithms 1 and 3, we obtain the following:

   $LCR(D_{bank.4.n.1})$ = {place, money, keep, pay, demand, activity },

   $CCR(D_{bank.4.n.1})$ = {money, pay, cash, capital, account, charge, ...
                     pay, bond, bill, charge, ...
                     money, cash, fund, check, ... } ,

   $LCR(E_{bank.4.n.1})$ = {rob(23), account(15), money(8), criminal(6), interest(5),
                     keep(5), paper(4), police(4), robber(4), thief(4),
                     cheque(3), ...}, and

   $CCR(E_{bank.4.n.1})$ = {officer, cop, detective, guard, protect, gangster, hoodlum,
                     larceny, hijacking, burglar, steal, fraud, swindle, …}.

Step 2: $WORD(bank,$ MONEY) =
                     $LCR(D_{bank.4.n.1}) + CCR(D_{bank.4.n.1}) + LCR(E_{bank.4.n.1}) + CCR(E_{bank.4.n.1})$.

   Similar calculations can be performed for other senses of *bank* to obtain *WORD*(*bank*, RIVER),
   WORD(*bank*, EARTH), etc.

Step 3: Compute *tf* and *idf* for each distinct word in *WORD*(*bank, S*).   For instance, there are 16 instances
   of *account* in *WORD*(*bank*, MONEY) and no other word list *WORD*(*bank, S*), S ≠ MONEY,
   contains *account*.   Thus, we have $tf_{account, MONEY}$ = 16 and $idf_{account}$ = 8.   Thus, the weight for
   account in *CR*(*bank*, MONEY) is $W_{account, MONEY}$ = 16 * 8 = 128.

Step 4: The weight $W_{X, S}$ in *CR*(*W, S*) for each word sense S is normalized to a sum of 100.   For instance,
   the total of the weights *CR*(*bank*, MONEY) = 6274.5; therefore, the normalized weight $W_{account, MONEY}$ = 2.04.   Table 8 shows more details about the contextual words and normalized weights in
   CR(*bank, S*) for all bank senses S.   The ten top-weighted context words from the CRs of the *bank*
   senses listed in Table 8 seem to be very relevant to each sense and to have strong collocates listed
   in BBI [Benson, Benson and Ilson 1993].   These weighted context words form the general CCR
   knowledge for senses of bank.   In the next section, we will show that this knowledge is effective
   for applying WSD to unrestricted text.

Table 8. *Top-ranked words in combined contextual representation based on definitions and examples of* bank *senses.*

Sense *S*	Context(with weights[4]) on *CR*(bank,*S*)
MONEY	rob(6.17), money(2.17), account(2.04), criminal(1.61), interest(1.40), keep(1.39), pay(1.18), police(1.07), robber(1.07), thief(1.07), …
RIVER	river(5.54), leave(2.18), towards(2.18), ship(1.23), city(1.09), dangerous(1.09), deer(1.09), descend(1.09), excavation(1.09), north(0.73), fish(0.56), …
EARTH	build(3.92), vole(3.92), earth(1.42), rise(0.98), flood(0.73), water(0.65), agricultural(0.20), barn(0.20), farm(0.20), garden(0.20), …
PILE	cloud(2.26), dark(1.97), heavy (0.99), storm(0.64), hall(0.36), shower(0.36), atmosphere(0.18), blizzard(0.18), blow(0.06), breeze(0.06), …
ROAD	moss(4.38), sit (2.19), wood (1.14), rest (1.09), gradient (0.18), junction (0.18), subway(0.08), tunnel(0.08), accelerator(0.06), accident(0.06), …
ROW	call(0.19), page(0.19), classical(0.16), compose(0.16), composition(0.16), leader(0.16), caravan(0.12), porter(0.12), bell(0.11), horn(0.11), …
MEDICINE	crutch(0.48), gut(0.48), abdomen(0.34), abdominal(0.34), ankle(0.34), anal(0.34), anus(0.34), aorta(0.34), pendicities(0.34), armpit(0.34), …
GAMBLE	club(0.43), cup(0.32), loser(0.24), win(0.24), defense(0.20), bet(0.17), champion(0.17), competition(0.17), gamble(0.17), games(0.17), …

3. Word Sense Disambiguating Algorithms

Among the recently proposed WSD systems, almost all have the property that the knowledge obtained is fixed when the system completes the training phase. This means that the acquired knowledge can not be enriched during the course of disambiguation. Such fixed knowledge is referred to as static knowledge. We believe that this property limits WSD performance. We propose lifting this limitation by adjusting the initial acquired knowledge to suit the text at hand. Alternatively, such expanded knowledge is referred to as adaptive knowledge. In this section, we will show how to distinguish between senses of text using adaptive disambiguation techniques. First, we will start with disambiguation of polysemous words in easy (trivial) contexts by using the fundamental knowledge previously acquired from MRD. Next, we will expand the acquired knowledge based on these disambiguated contexts. Finally, we will resolve the senses in the remaining contexts, called hard contexts.

3.1 Disambiguating Polysemous Words in Easy Contexts

The proposed WSD method starts with a simple disambiguation step using the topical CR described in a previous section. For instance, to disambiguate the word *bank* in Examples (4) through (6), the content words in its context are extracted, lemmatized and matched against the contextual representation of each of

---

[4] Weights for all contextual words of a sense are normalized to a sum of 100.

bank's word senses.   Each instance of *bank* is given a sense label in favor of a CR most similar to the context in question.   A sense label is assigned only when the match is strong enough and the runner-up sense is sufficiently weak.   In the following subsections, we will describe how to distinguish between strong and weak signals.   For instance, there is enough overlap between the CR for the instance of MONEY-bank in Examples (4) and (6) to warrant a sense label of MONEY-bank for the two instances of *bank*, but the match is not strong enough for the instance in Example (4).   We call Examples (4) and (6) easy[5] contexts, while Example (5) is a hard[6] context.

(4) … Participation *loans* are those made jointly by the SBA and *bank*s or other private *lending* institutions ...

(5) … individual action by every **nation** in position to help, we must squarely face this titanic challenge …

(6) … from *investment* firms all over the **nation**, all of them wanting a part of *shares* that would be *sold* (185,000 to the public at $12.50 with another 5,000 reserved for Morton Foods employers at $11.50 a *share*) there was even a cable in French from a *bank* in Switzerland that had somehow …

In addition, the contextual words closer to an ambiguous word may have greater influence on the sense of a word.   For instance, consider Example (7), where the intended sense of *bank* is MONEY. We observe that there are two salient words, *mortgage* and *river*, around an ambiguous word *bank*.   The word *mortgage* favors a MONEY sense, while the word *river* favors a RIVER sense.   Intuitively, the MONEY sense should be given more favorable consideration since *mortgage* is nearer to the ambiguous word than *river* is.   There are various representations for distance-based weights.   Here we adopt the metric proposed by Hawking and Thistlewaite [1995] to weigh the relevance of salient words in a text.

(7) … and an effort to get this religious center out of its rut of wild worship into a modern church organization. He emphasized to the Presiding Elder the plan of giving up the old church and moving across the river. The Presiding Elder was sure that that would be impossible. But he told Wilson to "go ahead and try". And Wilson tried. It did seem impossible. The *bank* which held the **mortgage** on the old church declared that the interest was considerably in arrears, and the real estate people said flatly that the land across the **river** was being held for an eventual development for white working people who were coming in, and that none would be sold to colored folk. When it was proposed to rebuild the church, Wilson found that the terms for …

To sum up, we outline a general WSD method using MRD-based contextual representation as Algorithm 4 for labeling an instance of a polysemous word *W* in a particular context *CON*(*W*).

---

[5] The algorithm for identifying easy contexts is Algorithm 4.

[6] The algorithm for resolving hard contexts is Algorithm 5.

Algorithm 4: (*StaticSense*) WSD using MRD-based contextual representation

Step 1: Preprocess the context and produce a list of lemmatized content words *CON(W)* in *W*'s context.

Step 2: For each sense S of W, compute the similarity between the context representation *CR(W, S)* and topical context *CON(W)*.

$$\text{Sim}\,(CR(W,S),\,CON(W)) = \frac{\sum\limits_{t \in M} (W_{t,s} + W_t)}{\sum\limits_{t \in CR(W,S)} W_{t,s} + \sum\limits_{t \in CON(W)} W_t}\,,$$

where $M = CR(W,S) \cap CON(W)$,

$W_{t,s}$ = the weight of a contextual word t with sense S in *CR(W)*,

W_t = the weight of t in *CON(W)* = $\dfrac{1}{\sqrt{|X_t|}}$,

X_t = the distance from *t* to *W* in number of words,

$\text{S*}(W, CON(W)) = \underset{s}{\arg\max}\ \text{Sim}\,(CR(W,S),\,CON(W)),$

$\text{S''}(W, CON(W)) = \underset{s}{\arg\max}\ \{\text{Sim}\,(CR(W,S),\,CON(W))\,|$

$\qquad\qquad \text{Sim}\,(CR(W,S),\,CON(W)) < \text{S*}(W, CON(W))\},$

$TSCORE(W, CON(W)) = \dfrac{\text{S*}(W,\ CON(W))}{\text{S''}(W,\ CON(W))}\,,$

$RANK\text{-}S(W, CON(W))$ = the rank of $\text{S*}(W, CON(W))$ among all $\text{S*}(X, CON(X))$ for all n instances of polysemous word X and context *CON(X)*,

$RANK\text{-}T(W, CON(W))$ = the rank of TSCORE(W, CON(W)) among TSCORE(X, CON(X)) for all n instances and context of polysemous word *X*.

Step 3: Construct the set of the triples T, where
T = { (*W, S, CON(W)*) | S = S*(*W, CON(W)*) such that
$RANK\text{-}S(W, CON(W)) \leq n/c$ and
$RANK\text{-}T(W, CON(W)) \leq n/c$,
where the constant $c \geq 1$ }[7].

Step 4: *DEFAULT(W)*= S such that the count of (*W, S, CON(W)*) ε T is the largest among all the senses of *W*.

Step 5: Assign (*W, CON(W)*) as the relevant sense S if (*W, S, CON(W)*) is in *T*, and assign *DEFAULT(W)* otherwise.

3.2 Adapting the Knowledge Base to Fit the Text

The adaptive approach to WSD hinges on two assumptions.   First, we assume that it is possible to build

---

[7] We use the WSD results of the top-ranking *c*'th instances in S* as well as TSCORE values which are more reliable. For instance, setting *c* to 2 amounts to taking the top-ranking 25% quantile of the test cases.

an initial general knowledge base so that a substantial portion of disambiguated text can be used to adapt the knowledge base to fit the text itself. The second condition for the adaptive approach to be feasible is that there is indeed new and effective information to be gained from the partially disambiguated text. In this section, we will first show the kinds of contexts in the Brown corpus and WSJ articles in which word sense ambiguity can be confidently resolved by using an MRD-based knowledge base. In these contexts, one will find an abundance of rich task-specific information not easily covered in a general or static knowledge base. We will also justify the use of contextual information and a task-specific default for WSD.

3.2.1 Discovering Task-specific Contextual Information

There is indeed an abundance of new and useful contextual information for word sense to be gained from typical, easy contexts. Such information can be extracted as long as ambiguity in these typical contexts can be interpreted successfully. For instance, the Brown corpus passage reproduced here as Example (8) is obviously very typical of MONEY-bank with salient words such as *accounts*, *stocks* and *property* in its context. Without a doubt, this instance of MONEY-bank can be resolved successfully using the kind of MRD-based knowledge base described in Section 3.1. Even though the overall context of this instance of MONEY-bank is a general one, it nevertheless contains many words, such as *law* and *state*, not in the MRD-based knowledge base. Such words might very well be incidental and have no intrinsic relation with the sense. For instance, the word *law* might just as likely be associated with RIVER-bank as MONEY-bank. Without much stretching of the imagination, it is possible to think of a likely event where *the state of Texas passes a law to declare an outer bank off limits to commercial development*. However, more often than not, these unexpected words will indicate real recurring contexts of word sense, either generally or in a task-specific way. Therefore, adapting the knowledge base to fit such a context is beneficial for WSD. For instance, the instances of *tree* and *camping* in the context of RIVER-bank in Example (9) seem to be reasonable additions to *CR*(*bank*, RIVER) in the sense that *tree* and *camping* are, in general, more strongly associated with RIVER-bank than with MONEY-bank. Even if that assertion generally does not hold, adding *tree* and *camping* to *CR*(*bank*, RIVER) as a way of adapting the knowledge base is still beneficial since it is likely to be valid in the very text where this association is discovered. The same argument holds for the local cue of *through* in the context of PILE-bank in Example (10), and for the instances of *donor* and *transfusion* in the context of MEDICINE-bank in Example (11). (See Table 9 for further details.)

> (8) … 63 million dollars at the end of the current fiscal year next Aug. 31. He told the committee the measure would merely provide means of enforcing the escheat law which has been on the books "since Texas was a republic". It permits the state to take over _bank_ **accounts**, **stocks** and other personal **property** of persons missing for seven years or more. The bill, which Daniel said he drafted personally, would force banks, insurance firms, pipeline companies and other …

(9) … On shooting preserves? Ask Sammy Shooter. WE WERE CAMPING a few weeks ago on Cape Hatteras Campground in that land of pirates, **seagulls** and **bluefish** on North Carolina's famed outer *bank*s. This **beach** campground with no *trees* or hills presents a constant **camping** show with all manner of equipment in actual use. With the whole camp exposed to view we could see the variety of canvas shelters in which Americans are camping now. There were …

(10) … to let down through the overcast and see the ground before it hit him. Bob Fogg didn't have today's advantages of Instrument Flight and Ground Control Approach systems. At the end of the calculated time he'd nose the Waco down **through** the **cloud** *bank* and hope to break through where some feature of the winter landscape would be recognizable. Usually back in Concord by noon, there was just time to get partially thawed out, refuel, and grab a bit of Mrs. Fogg's …

(11) … agreed, but explained that it would be necessary first to check Fred's blood to ascertain whether or not it was of the same type as Papa's. To give a **patient** the wrong type of blood, said the **doctor**, would likely kill him.   That was in the days before **blood** *bank*s, of course, and **transfusions** had to be given directly from *donor* to **patient**. One had to find a donor, and usually very quickly, whose blood corresponded with the patient's. And then it took considerably

Table 9. *Samples of disambiguated topical contexts of* bank *in the Brown corpus.*

Sense	Example No.	Typical, Easy Context of Various Senses of *bank*	General Topical Context	Task-specific Context
MONEY	(9)	… It permits the state to take over *bank* **accounts**, **stocks** and other personal **property** of persons missing for seven years or more. …	Account Stock Property	law bill
RIVER	(10)	… WE WERE CAMPING a few weeks ago on Cape Hatteras Campground in that land of pirates, **seagulls** and **bluefish** on North Carolina's famed outer *bank*s. …	Seagull Bluefish Hill	tree camping
PILE	(11)	… At the end of the calculated time he'd nose the Waco down *through* the **cloud** *bank* and hope to breakthrough where some feature of the winter landscape would be recognizable. …	Cloud	flight through
MEDICINE	(12)	… That was in the days before **blood** *bank*s, of course, and *transfusions* had to be given directly from *donor* to **patient**. …	blood doctor patient	donor transfusion

3.2.2 Using the Default Sense

The distribution of senses of a word might not follow Zipf's law because their rank-frequency plot does not follow the power-law well, and it is often quite skewed even in a balanced corpus.   In the Brown corpus,

60% of the instances of twelve polysemous words are the top-ranking sense of the word, according to an experimental report by Luk [1995].

Generally, the top-ranking sense of a word is corpus-dependent. Table 10 presents some statistics about the distribution of senses in different corpora. For instance, we find that CURIOSITY-interest is favored over MONEY-interest 194 to 49 in the Brown corpus, while preference is reversed with counts of 53 and 122 in the WSJ corpus. On the other case, GRAMMAR-sentence is favored over JUDGEMENT-sentence 22 and 10 in the Brown corpus while preference is reversed with counts of 1 to 11 in the WSJ corpus. Using a fixed default would be disastrous for *interest* or *sentence* in at least one of these corpora. The adaptive method alternatively uses a set of disambiguated samples from the text in question to estimate the default.

Table 10. *Skewed sense distribution is corpus dependent*

Word	Sense	Brown	WSJ
Interest	MONEY	49	122
	CURIOSITY	194	53
Sentence	GAMMAR	22	1
	JUDGEMENT	10	11
Bass	MUSIC	15	2
	FISH	1	0

3.3 The Adaptive WSD Algorithm

We are now ready to present a new adaptive approach to WSD based on the fundamental knowledge base acquired from MRD. Previous sections have already shown how such a knowledge base can be built and described its advantages. We will show one way of using a MRD-based knowledge base for WSD. Although the knowledge base does not guarantee high precision and 100% coverage, a substantial portion, say 50%, can be disambiguated at a high precision rate. In this section, we will show how such a level of coverage and high precision can be put to use in an adaptive way to maintain the same high precision rate at 100% coverage. We will first describe the adaptive algorithm. Examples will be given in Section 3.4 to illustrate how the algorithm works and to give some idea of the potential effectiveness of adaptation.

The algorithm starts with an initial disambiguation step using the knowledge base derived from the MRD. An adaptation step follows which produces a knowledge base from the partially disambiguated text. Finally, the undisambiguated part is disambiguated according to the adapted knowledge base. Algorithm 5 gives a formal and detailed description.

Algorithm 5: (*AdaptSense*) Adaptive WSD

Step 1: Run Algorithm 4 to obtain triples T_1 of word, word sense and context.

Step 2: From the selected triples $(W, S, CON(W)) \in T_1$, compute a new set of contextual representations:

$WORD(W,S) = \{ u \mid u \in CON(W) \text{ and } (W, S, CON(W)) \in T_1 \}$.

Step 3: Build the contextual representation $CR(W,S)$ of sense S of word W from $WORD(W, S)$ according to Algorithm 3.

$DEFAULT(W) = S$ such that the count of $(W, S, CON(W)) \in T_1$ is the highest among all the senses of W.

Step 4: For all the instances of polysemous W and its $CON(W)$ such that $(W, S, CON(W))$ is not in T_1 (for all senses S of W),

$$\text{Sim}\,(CR(W,S), CON(W)) = \frac{\sum\limits_{t \text{ in } M} (W_{t,s} + W_t)}{\sum\limits_{t \text{ in } CR(W,S)} W_{t,s} + \sum\limits_{t \text{ in } CON(W)} W_t}\,,$$

where $\quad M = CR(W,S) \cap CON(W)$,

$W_{t,s} =$ the weight of a contextual word t with sense s in $CR(W,S)$,

$W_t =$ the weight of t in $CON(W) = \dfrac{1}{\sqrt{|X_t|}}$,

$X_t =$ the distance from t to W in number of words.

$S^*(W, CON(W)) = \underset{s}{\arg\max}\,\text{Sim}\,(CR(W,S), CON(W))$,

$S''(W, CON(W)) = \underset{s}{\arg\max}\,\{\,\text{Sim}\,(CR(W,S), CON(W)) \mid$

$\text{Sim}\,(CR(W,S), CON(W)) < S^*(W, CON(W))\,\}$,

$TSCORE(W, CON(W)) = \dfrac{S^*(W,\ CON(W))}{S''(W,\ CON(W))}\,,$

$RANK\text{-}S(W, CON(W)) =$ the rank of $S^*(W, CON(W))$ among all

$S^*(X, CON(X))$ for all n instances of polysemous

word X and context $CON(X)$,

$RANK\text{-}T(W, CON(W)) =$ the rank of $TSCORE(W, CON(W))$ among

$TSCORE(X, CON(X))$ for all n instances and context of polysemous word X.

Step 5: Construct the set of triples T_2, where

$T_2 = \{ (W, S, CON(W)) \mid S = S^*(W, CON(W)) \text{ such that}$

$RANK\text{-}S(W, CON(W)) \leq n/c$ and

$RANK\text{-}T(W, CON(W)) \leq n/c$,

where the constant $c \geq 1 \}$.

Step 6:   Assign (*W*, *CON*(*W*)) to the relevant sense *S*, such that

$$(W, S, CON(W)) \in T_1, \text{ or}$$

$$(W, S, CON(W)) \in T_2, \text{ or}$$

$$DEFAULT(W), \text{ otherwise.}$$

3.4 An Illustrative Example

To show how Algorithm 5 works in an adaptive fashion, we will consider the case of disambiguating the Brown corpus, focusing on the polysemous word *bank*.   For this purpose, we will describe step by step how the algorithm operates on the two following passages in the Brown corpus containing an instance of *bank*.   The two passages are reproduced here as Examples (12) and (13), showing a context window of 50 words before and after the polysemous word which is used in the algorithm for disambiguation.

(12)   … to face charges of assault and robbery, Portland detectives said Friday. Mrs. Lavaughn Huntley is accused of driving the getaway car used in a robbery of the Woodyard Bros' Grocery, 2825 E. Burnside St., in April of 1959. Her husband, who was sentenced to 15 years in the federal prison at McNeil Island last April for robbery of the hillsdale branch of Multnomah *Bank*, also was charged with the store holdup. Secret Grand Jury indictments were returned against the pair last week, Detective Murray Logan reported. The Phoenix arrest culminates more than a year's investigation by Detective William Taylor and other officers. Taylor said Mrs. Huntley and her husband also will be questioned about …

(13)   … Of cattle in a pasture without throwin' 'em together for the purpose was called a "pasture count". The counters rode through the pasture countin' each bunch of grazin' cattle, and drifted it back so that it didn't get mixed with the uncounted cattle ahead. This method of countin' was usually done at the request, and in the presence, of a representative of the *bank* that held the papers against the herd. The notes and mortgages were spoken of as "cattle paper". A "book count" was the sellin' of cattle by the books, commonly resorted to in the early days, sometimes much to the profit of the seller. This led to the famous sayin' in the Northwest of the "books won't freeze". This became a common byword durin' the …

Step1: Identifying an easy context

This step corresponds to five substeps of Algorithm 4.   First, only the salient words that are in *CR*(*bank*, *S*) for *S* in {MONEY, RIVER, EARTH, PILE, ROW, ROAD, MEDICINE, GAMBLE} are of interest; all other words are thrown out for now.   To calculate similarity values, the weights for these words with respect to relevant senses are pulled out from the initial knowledge base.   Tables 11 (a) and (b) show these words, their position relative to *bank*, and their weights according to a knowledge base extracted from LDOCE.

Table 11(a). *Weights for salient words in Example* (*12*) *for* bank *in the initial WSD stage.*

	X_t	W_t	$W_{t,s}$ in CR (bank, S)							
			S_{MONEY}	S_{RIVER}	S_{EARTH}	S_{PILE}	S_{ROW}	S_{ROAD}	$S_{MEDICINE}$	S_{GAMBLE}
drive	-46	0.15	-	0.81	0.02	0.02	0.07	0.01	-	0.04
getaway	-44	0.15	1.37	-	-	-	-	-	-	-
car	-43	0.15	0.94	-	-	0.81	0.12	0.07	-	-
robbery	-39	0.16	0.81	-	-	-	-	-	-	-
husband	-24	0.20	1.07	-	-	-	-	-	-	-
year	-18	0.24	0.81	-	-	-	-	-	-	-
prison	-14	0.27	2.04	-	-	-	-	-	-	-
robbery	-7	0.38	2.31	-	-	-	-	-	-	-
branch	-3	0.58	3.81	-	-	-	-	-	-	-
bank	0									
charge	3	0.58	3.58	-	-	-	-	-	-	-
return	13	0.28	1.77	-	-	-	-	-	-	-
week	18	0.24	1.34	-	-	-	-	-	-	-
report	22	0.21	-	-	-	-	-	0.90	-	-
year	30	0.18	0.81	-	-	-	-	-	-	-
husband	45	0.15	1.07	-	-	-	-	-	-	-

The context of Example (12) resembles the CR of MONEY-bank the most. Table 11(a) indicates clearly that very salient words in *CR*(*bank*, MONEY-bank), such as *robbery*, *branch*, and *charge*, occur in close proximity to the word *bank*. Although words related to other senses, such as *drive* and *report*, do occur, they are fewer and are located at quite a longer distance. It is not surprising that the similarity of this context with MONEY-bank and the *t*-score ranks high enough for this instance to be included in T_1.

On the other hand, Example (13) does not resemble the CR of any particular sense of *bank* more than it does those of other senses. That is evident from the weights shown in Table 11(b). The only indicative word *representative* is not enough to enable interpretation of the intended sense of MONEY-bank. All other words are either not in any CRs or ambivalent (*hold*, *note* and *paper*), indicating a number of senses competing with MONEY-bank. Hence, the similarity of this context with MONEY-bank and the *t*-score do not rank high enough for this instance to be included in T_1.

Table 11(b). *Weights for salient words in Example (13) for* bank *in the initial WSD stage.*

Word	X_t	W_t	$W_{t,s}$ in CR (bank, S)							
			S_{MONEY}	S_{RIVER}	S_{EARTH}	S_{PILE}	S_{ROW}	S_{ROAD}	$S_{MEDICINE}$	S_{GAMBLE}
call	-50	0.14	-	-	-	-	-	0.99	-	-
pasture	-48	0.14	-	-	1.00	-	-	-	-	-
count	-47	0.15	0.82	-	-	-	-	-	-	0.86
counter	-45	0.15	-	-	-	-	-	-	-	0.92
ride	-44	0.15	-	-	-	-	-	-	-	0.97
pasture	-41	0.16	-	-	1.00	-	-	-	-	-
ahead	-20	0.22	-	0.94	-	-	-	-	-	-
representative	-3	0.58	3.04	-	-	-	-	-	-	-
bank	0									
hold	2	0.71	-	-	3.65	-	-	3.03	-	3.04
paper	4	0.50	1.07	0.81	-	0.81	-	0.82	-	-
note	9	0.33	1.79	-	-	-	-	1.58	-	-
paper	17	0.24	1.07	0.81	-	0.81	-	0.82	-	-
book	19	0.23	0.93	-	-	-	-	0.89	-	-
count	20	0.22	0.82	-	-	-	-		-	0.86
book	28	0.19	0.93	-	-	-	-	0.89	-	-
profit	40	0.16	0.84	-	-	-	-		-	-
lead	45	0.15	-	0.81	-	0.82	0.92	0.89	-	-

Step 2: Computing a new contextual representation set based on an easy context

With the instance *bank* in Example (12) resolved to MONEY-bank, the following triple is created and added to T_1.

> (*bank*, MONEY-bank, "to face charges of assault and robbery, Portland detectives said Friday. Mrs. Lavaughn Huntley is accused of driving the getaway car used in a robbery of the Woodyard Bros' Grocery, 2825 E. Burnside St., in April of 1959. Her husband, who was sentenced to 15 years in the federal prison at McNeil Island last april for robbery of the Hillsdale branch of Multnomah **Bank**, also was charged with the store holdup. Secret Grand Jury indictments were returned against the pair last week, Detective Murray Logan reported. The Phoenix arrest culminates more than a year's investigation by Detective William Taylor and other officers. Taylor said Mrs. Huntley and her husband also will be questioned about")

From the triples T_1, a list *WORD*(S) of contextual words for each sense S of word *bank* and the most frequent sense *DEFAULT*(*bank*) are calculated. Therefore, the contextual words in the triple from Example (12) will be lemmatized. With stop words removed, we obtain a list like the following:

WORD(MONEY-bank) = {face, charge, assault, robbery, portland, detectives, say, Friday, mrs, lavaughn, huntley, accuse, drive, getaway, car, use, robbery, woodyard, bros, grocery, burnside, st, april, husband, sentence, year, federal, prison, mcneil, island, last, april, robbery, hillsdale, branch, multnomah, charge, store, holdup, secret, grand, jury, indictment, return, against, pair, last, week, detective, murray, logan, report, phoenix, arrest, culminate, year, investigation, detective, william, taylor, officer, taylor, say, mrs, huntley, husband, question, … }

Step 3: Assigning weight to the contextual representation

From the word lists for all senses of *bank*, the new set of CRs can be derived.   The *CR*(*bank*, S) for the word sense *S* basically consists of every word in *WORD*(*S*) associated with a weight.   Weights are assigned in favor of contextual words frequently occurring in the context of a particular word sense and that of a smaller number of other senses.   For instance, the word *cooperative* occurs very frequently and only in the context of MONEY-bank in the part of the Brown corpus resolved in Step 1.   Table 12 (b) shows a completely new set of CRs for *bank*'s senses which are obviously quite different from the LDOCE-based CR shown in Table 12(a).   Intuitively, this set of CRs should be more relevant to the Brown corpus than the MRD-based ones.

According to our experiment, there are quite a number of *bank* instances in the Brown corpus that are very typical and can be reliably resolved using LDOCE-based contextual representation.   Those instances are predominately resolved as MONEY-bank.   Therefore, we have *DEFAULT*(*bank*) = MONEY-bank.

Table 12 (*a*). *Selected sample of initial knowledge for* bank *senses.*

Sense	Top-ranking Contextual Words (with weights)
MONEY	rob(6.17), money(2.17), account(2.04), interest(1.40), pay(1.17), robber(1.07), month(0.80), robbery(0.81), prison(0.81), year(0.81), charge(0.58), …
RIVER	river(5.54), ship(1.23), deer(1.09), hunter(1.09), drive(0.81), fish(0.55), air(0.54), hill(0.44), east(0.36), south(0.36), boat(0.13), boatman(0.13), …
EARTH	build(3.91), water(0.65), sky(0.29), plant(0.19), west(0.17), north(0.11), south(0.11), sidewalk(0.02), street(0.02), drive(0.02) , bridge(0.01), …
PILE	wet(0.29), basin(0.06), window(0.06), table(0.03), clay(0.02), cloth(0.02), drive(0.02), …
ROAD	moss(4.38), sit(2.19), wood(1.14), subway(0.18), car(0.07), drive(0.01), …
ROW	car(0.12), letter(0.09), write(0.09), visit(0.08), drive(0.07), column(0.04), story(0.04), …
MEDICINE	blood(0.33), body(0.33), shoulder(0.33), patient(0.14), doctor(0.07), course(0.03), …
GAMBLE	box(0.11), pocket(0.11), play(0.08), check(0.05), point(0.05), drive(0.04), …

Steps 4-6: Disambiguating a hard context

Armed with the new CRs, the instances that do not pass the test in Step 1 are re-evaluated again in Step 4. The similarity for each of those instances, including Example (13), is re-calculated for all possible word senses.   The new weights for contextual words in Example (13) are shown in Table 13.

Contrary to the situation in Step 1, where the MRD-based CR is used, there are now more words in the context that are indicative of the intended sense.   From the perspective of the new CRs, the words *method*, *usual*, *request*, *paper*, *note*, and *book* all point to the sense of MONEY-bank and not to any other sense.   These words either do not exist or ambivalent with respect to the MRD-based CR.   As a whole, these words provide enough evidence to reverse the previous inconclusive situation leading to the expected sense of MONEY-bank.   In the event that the maximal similarity is lower than a threshold value, the default sense of MONEY-bank is used.   In this particular case, the default happens to be correct.

Table 12 (*b*). *Selected sample of adaptive knowledge for* bank *senses.*

Sense	Selected Contextual Words (with weights)
MONEY	cooperatives(1.50), department(1.37), affairs(1.07), export-import(0.69), federal(0.60), government(0.54), short-term(0.43), cooperative(0.41), administration(0.38), firm(0.35), sponsor(0.35), …
RIVER	church(0.80), soldier(0.68), dill(0.66), camping(0.54), fame(0.52), outer(0.52), rhine(0.52), motel(0.40), sight(0.40), tree(0.40), camp(0.33), …
EARTH	manchester(4.35), company(3.77), telegraph(3.77), goodwin(3.11), power(1.88), light(1.69), door(1.64), cemetery(1.31), commercial(1.31), dwelling(1.31), electric(1.31), business(1.23), construction(1.23), …
PILE	tiber(3.69), fold(2.83), moonlight(1.84), thick(1.84), anatomy(0.98), bedside(0.98), buckle(0.98), damn(0.98), dancer(0.98), dark(0.58), …
ROAD	-
ROW	feel(8.51), error(5.37), correct(3.58), shareholder(3.47), people(3.36), data(0.89), fund(0.89), funds(0.89), …
MEDICINE	stumbled(4.51), transfusions(3.46), donor(2.25), frail(2.25), child(1.20), laboratory(1.20), neck(1.20), night(1.20), sample(1.20), …
GAMBLE	fraud(4.10), drink(2.85), grade(2.67), stare(2.67), chief(1.42), collusion(1.42), conclusive(1.42), death(1.42), …

4. Experiment and Evaluation

4.1 Experiment

The experimental setup can be described in a number of steps as follows.   (1) A set of 13 polysemous words was selected as the target for disambiguation and evaluation.   (2) For each of the polysemous

words, a sense division was established based on the LDOCE treatment of relevant nominal senses.　The LDOCE's sense division was used largely as is, with only a couple of closely related senses merged.　(3) Two sets of text from corpora were gathered as the *test sets*.　(4) Two human judges were asked to assign a sense label to each nominal instance of these 13 words in the two test sets.　(5) Two WSD programs were written to disambiguate nominal instances of these polysemous words in the test sets.　(6) The results of running the two programs on both test sets were compared against those of human assessors. The number of test instances and that of correctly disambiguated ones in these four experiments were tallied to produce a precision rate for each experiment.　In the following, we describe each step in turn.

Table 13. *Weights for salient words in Example* (*13*) *after adaptation.*

Word	X_t	W_t	S_{MONEY}	S_{RIVER}	S_{EARTH}	S_{PILE}	S_{ROW}	S_{ROAD}	$S_{MEDICINE}$	S_{GAMBLE}
			$W_{t,s}$ in CR(bank, S)							
cattle	-21	0.22	0.83	1.00	-	-	-	-	-	-
ahead	-20	0.22	-	1.06	-	-	-	-	-	-
method	-18	0.24	0.86	-	-	-	-	-	-	-
usual	-14	0.27	1.53	-	-	-	-	-	2.10	-
request	-10	0.32	1.62	-	-	-	-	-	-	-
representative	-3	0.58	3.06	-	-	-	-	-	-	-
bank	0									
hold	2	0.71	3.01	3.31	-	3.33	-	-	-	-
paper	4	0.50	0.95		-	-	-	-	-	-
herd	7	0.38	-	1.76	-	-	-	-	-	-
note	9	0.33	1.59		-	-	-	-	-	-
mortgage	11	0.30	1.54	1.63	-	-	-	-	-	-
speak	13	0.28	-	1.64	-	-	-	-	-	-
cattle	16	0.25	0.83	1.00	-	-	-	-	-	-
paper	17	0.24	0.95	-	-	-	-	-	-	-
book	19	0.23	0.89	-	-	-	-	-	-	-
cattle	25	0.20	0.83	1.00	-	-	-	-	-	-

(1) Test words

We limited our experiment and evaluation to a set of thirteen words with higher than usual ambiguity. That is due mainly to the fact that the process of evaluation is a difficult and expensive one.　It is often difficult to pin down the number of senses allowed for a word in the experiment.　For the purpose of comparing results with other approaches, we stick to words that have been studied in various experiments reported in the literature on computational linguistics.　These words include *bank*, *bass*, *bow*, *cone*, *duty*, *galley*, *interest*, *issue*, *mole*, *sentence*, *slug*, *star*, and *taste*.

(2) Sense division

The sense division for each of these test words was very crucial in the WSD experiment. We used a sense division based on LDOCE's treatment of the nominal senses of these words. The division is somehow more fine-grained than those used in other WSD studies. This level of sense division is very close to the kind of granularity required for machine translation. For most cases, a word sense has a unique Chinese translation.

(3) Test corpora

We aimed to determine the effectiveness of the proposed approach for unrestricted text and to find out how domain and genre affect WSD. Therefore, we used the Brown corpus and a collection of WSJ articles from October 30 to November 2, 1989 as the test sets. Passages of 100 words centered at an instance of the test words in the two corpora were extracted using a SED program. It is in general not hard to write a regular expression in the SED program to exclude verbal instances, so only a small number of verb cases were extracted. These verbal instances were excluded from the experiment according to the marks made by human judges. For these thirteen words under investigation, we had 846 and 903 passages of nominal senses from the Brown corpus and WSJ test sets, respectively.

(4) Judgement

To be as subjective as possible, we asked two human judges to assign a sense label to each nominal instance of these thirteen words in the two test sets. There were also cases which fell out of the scope of our sense division. Most of these cases used proper nouns, so they bore none of the meaning represented in our sense division. Cases judged to be verbal uses or proper names were removed from the test cases. For instance, the word *bow* in a Brown corpus passage, reproduced here as Example (14), was an instance of a proper name and, therefore, was excluded from the test cases.

(14)   … The announcement that the secrets of the Dreadnought had been stolen was made in *Bow* St. police court here at the end of a three day hearing …

(5) Static vs. Adaptive WSD

In the previous sections, we argued in favor of using an MRD-derived knowledge base because we believe that the fundamental information in an MRD can be very helpful for WSD. Despite our belief in the effectiveness of the MRD-derived knowledge base, we also expected that adaptation could improve its effectiveness a bit further. Therefore, we implemented programs for both Algorithms 4 and 5. These two programs were executed in order to disambiguate the test cases in the Brown and WSJ corpora.

4.2 Evaluation

The results of running the two programs on both test sets were compared against those of human assessors. The number of test instances and that of correct assignments in these four experiments were tallied to

calculate the precision rate for each experiment. All results were based on 100% applicability[8]. Statistics for the experimental results are summarized in Tables 14 and 15. Several observations can be made based on the results. First, evidently, the MRD-based knowledge base was reasonably helpful for WSD. The results shown in Tables 14 and 15 indicate that without adaptation, the knowledge extracted from LDOCE and LLOCE could be used to deliver a precision rate of 65.2% for the Brown corpus and 76.6% for the WSJ articles. Second, adaptation indeed helped boost the precision rate by over 5% for the Brown corpus. As for the WSJ test set (see Table 15), adaptation only marginally increased the average precision rate. Closer examination of the results for this test set shows that three words, *bank*, *interest*, and *issue*, dominated the experiment and evaluation results. The precision rate for *bank* was over 95%, which left adaptation with very little room for improvement. The other two words, *interest* and *issue*, were very general and difficult to disambiguate.

5. Discussion

Although it is often difficult to compare results from experiments based on different domains, genres and setups, the experimental results presented here seem to compare favorably with the experimental results reported in previous WSD research. Our adaptive approach could disambiguate with an average precision rate of 71.2% for these thirteen words in Brown and of 76.5% for these words in WSJ. For the Brown corpus, Luk [1995] experimented with the same words we used except for the word *bank* and reported that there were totally 616 instances of these words (slightly less than the 749 instances we found). The precision rate for all instances was 60%. Leacock, Towell and Voorhees [1993] reported a precision rate of 76% for disambiguating the word *line* in a sample of WSJ articles.

---

[8]Applicability (coverage) denotes the proportion of cases in which the WSD model performed disambiguation.

Table 14. *Disambiguation results for thirteen ambiguous words in the Brown corpus.*

Word	Sample Sizes	StaticSense	AdaptSense	
		# of correct	# of correct in 1st run	# of correct in 2nd run
bank	97	68	71	71
bass	16	16	16	16
bow	12	3	3	2
cone	14	14	14	14
duty	75	67	69	69
galley	4	4	4	4
interest	346	213	228	226
issue	141	67	88	97
mole	4	2	2	2
sentence	32	30	30	30
slug	8	4	6	6
star	46	28	29	29
taste	51	36	36	36
Total	846	552	596	602
precision		65.2%	70.5%	71.2%

Besides the precision rate, a number of interesting features of this approach are also important. First, the proposed disambiguation system is robust and portable, since absolutely no corpus-specific knowledge is needed in the disambiguation procedure. It can be applied readily to test data in a variety of domains and genres with performance rivaling that of methods requiring a substantial training corpus. Second, the proposed approach is considerably more time efficient when compared to other learning strategies. Although the bootstrap approach proposed by Yarowsky [1995] has an element of adaptation to it, his method still requires a long training process to derive a static knowledge base for WSD. The differences between our method and his lie in the initial knowledge, the level of abstraction, and the learning cycle. We propose to exploit rich conceptualized knowledge from MRD at the outset, while the bootstrap method uses merely a couple of word collocations for each sense to start the learning process. Since the bootstrap method aims to derive a word-based conceptual representation with a large parameter space, a very large training corpus is required. The thesaurus used in the proposed approach provides an appropriate level of abstraction and, thus, alleviates the need for a very large corpus. The time required for learning in the two approaches is also quite different. The adaptive approach requires a single round of adaptation for effective WSD, while the bootstrap method needs many rounds of learning.

Table 15. *Disambiguation results for thirteen ambiguous words in the WSJ test set.*

Word	Sample Sizes	StaticSense	AdaptSense	
		# of correct	# of correct in 1st run	# of correct in 2nd run
Bank	370	350	353	353
Bass	2	2	2	2
Bow	-	-	-	-
Cone	-	-	-	-
duty	25	19	22	22
galley	-	-	-	-
interest	221	123	127	122
issue	260	181	177	175
mole	-	-	-	-
sentence	12	11	12	12
slug	-	-	-	-
star	7	3	2	2
taste	6	3	3	3
Total	903	692	698	691
precision		76.6%	77.3%	76.5%

Speedy adaptation is the consequence of using rich conceptualized knowledge to start the learning process. To show that this is truly the case, we have revised Algorithm 5 by adding a second and a third adaptation step and by applying the new CR to a reserved batch of low-ranking instances instead of using defaults. The results obtained using more adaptation steps are shown in Figure 2. The precision rates show that the additional adaptation steps have only a marginal effect.
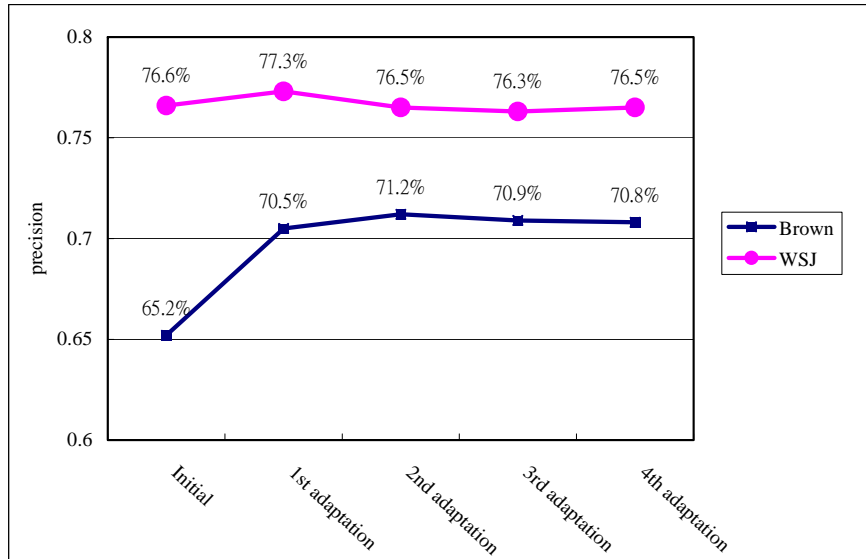
Figure 2 *Average precision rates with and without adaptation.*

One of the limiting factors of this approach is the quality of sense definition in the MRD. Short and vague definitions tend to lead to inclusion of inappropriate topics in the contextual representation. With such inferior CRs, it is not possible to produce enough precise samples in the initial step for subsequent adaptation. For instance, it is difficult to derive appropriate contextual knowledge for the LDOCE senses in (15) since their definitions mainly consist of either function words or very common words:

(15) **interest.1.n.1**   a readiness to give attention

　　　issue.1.n.1   the act of coming out

　　　issue.1.n.2   an example of this

　　　issue.1.n.3   something which comes or is given out

　　　issue.1.n.4   the act of bringing out something in a new form

The experiment and evaluation results show that adaptation is most effective when a high-frequency word with contrasting senses is involved. For low-frequency senses, such as EARTH, ROW, and ROAD senses of *bank*, the approach does not seem to be very effective. That is not a problem specific to the adaptive approach, and all other approaches in the literature suffer from the same problem of data sparseness. Even with static knowledge acquired from a very large corpus, these senses were disambiguated at a considerably lower precision rate than other senses.

6 Related Work

There has been increasing interest in using a machine to identify the intended sense of a polysemous word in a given context. Recently, various approaches to WSD have been proposed in the natural language processing literature, and old ideas have been superseded by newer ones at a rapid rate. Central to these development efforts are the kind of contextual knowledge encoded and the way this knowledge is represented and acquired. In this section, we review the recent literature on WSD from the perspectives of different types of contextual knowledge and their representational schemes.

6.1 Lexicalized vs. Conceptual Encoding of Context

Any kind of scheme for acquiring contextual information of word sense must begin with a way of identifying the word sense since word sense is an abstract concept not clear on the surface. Once this is done, we can use the surrounding words to build a contextual representation of the word sense for WSD. There are three approaches to the chicken-and-egg problem of dividing word senses. First, one can resort to human intervention to get a hand-tagged corpus of word senses. Most early WSD works used this approach and went to the trouble of hand-tagging the intended sense of each polysemous word in the training corpus [Kelly and Stone 1975; Hearst 1991]. Second, one can take the numbered sense entries readily available in a machine-readable dictionary and treat their definitions and examples as contextual information [Lesk 1986; Veronis and Ide 1990; Wilks *et al.* 1990; Guthrie *et al.* 1991]. The third way of identifying word sense exploits linguistic constraints. For instance, three linguistic constraints can be exploited for successful sense tagging and WSD.

• **One sense per discourse** The senses of all instances of a polysemous word are highly consistent within any given document.

•**One sense per collocation** Nearby words provide strong and consistent clues to the sense of a target word, conditional on the relative distance, order, and syntactic relationship.

•**One sense per translation** Translations in a bilingual corpus can be used to represent the senses of words.

As an example of the first constraint, consider the word *suit*. The constraint captures the intuition that if the first occurrence of *suit* is a LAWSUIT sense, then later occurrences in the same discourse are also likely to refer to LAWSUIT [Gale, Church and Yarowsky 1992a]. The second constraint indicates that most works on statistical disambiguation have made the basic assumption that word sense is strongly correlated with certain contextual features, like occurrence of particular words in a window around the ambiguous word. However, Yarowsky [1995] proposed an approach in which strong collocations were identified for WSD. If a bilingual corpus was available, differences in translations of the polysemous word allowed one to delineate the intended sense, particularly in the case of contrasting polysemy. Gale,

Church and Yarowsky [1992b] used French translations in parallel texts to disambiguate some polysemous words in English. For instance, the senses of *duty* were usually translated as two different French words, *droit* and *devoir*, respectively, representing the senses *tax* and *obligation*. Thus, a number of *tax* sense instances of *duty* could be collected by extracting instances of *duty* that were translated as *droit,* and the same could be done for *obligation* sense instances of *duty*.

Once word senses are identified in one way or another, the context of a particular word sense can then be acquired and encoded in some way for use in the subsequent disambiguation step. There are at least two ways of encoding contextual knowledge. The obvious way, the lexicalized representation, is a surface scheme that keeps a weighted list of words appearing in the context of a particular sense. On the other hand, the conceptual representation encodes the classes of words that might appear in the context.

6.1.1 Lexicalized Representation of Context

Dictionary Definitions as Context Lesk [1986] described a word-sense disambiguation technique based on the number of overlaps between words in a dictionary definition and the fixed-size window of words surrounding the target. The author reported WSD performance ranging from 50% to 70% when the method was applied to a sample of ambiguous words. Lesk's method had failed to determine the correct senses of words when two or more senses of a word had the same number of overlaps with the context. Veronis and Ide [1990] constructed an artificial neural network from sense definitions, representing each word in the definition text as a node in the network. Different senses of each word competed with each other through the mechanism of spreading activation initiated at the nodes of contextual words. White [1988], Guthrie *et al.* [1991], and Slator [1991] used measures of words in context overlapping with dictionary definitions. One major problem of these earlier approaches was their lack of abstraction. The rich semantic information in the definition, such as the genus term, differentia, and implicit topics, was not exploited to the fullest.

Context as Co-occurrence Probabilities Gale, Church and Yarowsky [1992b] indicated that translation in a bilingual corpus could be used to provide tagged material for supervised learning of WSD knowledge. In their experiment, French translations were, in effect, used to represent the senses of some English words under the assumption of *one-sense-per-translation*. The Bayesian model was used to represent the contextual words in terms of their probabilities of occurrence. They reported a 90% accuracy rate in discriminating between two contrasting senses of six ambiguous nouns in the Canadian Hansards: *duty, drug, land, language, position, and sentence*. The weaknesses of this approach include the dreaded problem of data sparseness. Even when a very large corpus is available, it is still difficult to guarantee that each word sense will have enough contextual samples to avoid running into the problem of zero frequency, namely, the difficulty of assigning appropriate probabilistic values to words that do not appear in these contextual samples.

Smoothing Co-occurrence Probabilities   Yarowsky [1992] improved on the WSD method proposed by Gale, Church and Yarowsky [1992b] by smoothing the concurrence probability via predefined semantic classification.   Basically, that was done by lumping the probabilities related to all the senses in a thesaurus category to smooth the zero frequency cases.   For instance, the contextual information of *bird* and other animals was used to build a contextual representation for all the senses in the animal category in Roget's Thesaurus [1987].   His experiment showed in a *close test* using Grolier's Encyclopedia that instances of twelve words, *bass, bow, cone, duty, galley, interest, issue, mole, sentence, slug, star,* and *taste,* could be disambiguated with an average precision rate of 92%.   However, a very large corpus is required to train such a lexicalized contextual model, and clearly this kind of static model has a portability problem.

6.1.2 Conceptual Representation of Context

Context as Definition-Based Conceptual Co-occurrence   Luk [1995] advocated using defining words in the MRD for the contextual representation of word sense.   Reminiscent of an earlier work by Wilks *et al.* [1990], Luk proposed a definition-based concept co-occurrence model (DBCC) for WSD.   With the model, the context of each word sense is represented using a vector of LDOCE defining words in the sense definition.   The author argued that by using a fixed, relatively small number of concepts, a small corpus could provide enough concept co-occurrence data for statistical sense disambiguation.   In a close test, the DBCC model trained on the Brown corpus was found to be capable of disambiguating 60%[9] of the instances of the same twelve ambiguous words used in Yarowsky's experiment.

Context as Thesaurus Categories   Many researchers have exploited the semantic categories in a thesaurus, such as Roget's and LLOCE, or the subject information in a dictionary for context representation and WSD.   Walker and Amsler [1986] applied subject codes in LDOCE as semantic representation for WSD.   Black [1988] reported an accuracy rate of around 50% when Walker and Amsler's algorithm was applied to a sample of five ambiguous words: *interest, point, power, state,* and *term*.   Pure conceptual representation is the most economical kind of WSD model since it requires the smallest parameter space and requires no substantial texts for training.   Chen *et al.* [1996] proposed a mixed representational scheme for context based on contextual words as well as LLOCE topics.   With a contextual representation acquired from example sentences in LDOCE/E-C, the authors reported that the method could disambiguate around 70% of the instances of thirteen polysemous words in the Brown corpus.

---

[9] The originally reported value, 77%, was based on the average of the precision rates for all twelve words.   This form of evaluation is sensitive to the outcome of a handful of test samples since the precision rate of a word with a couple of samples could have an overly strong impact on the average.   In this paper, we use the average rate of precision calculated over all instances.

6.2 Topical vs. Local Representation of Context

In almost all the studies described in Section 2.1, topical context was used in WSD. In a number of research works related to machine translation, researchers have used local context to solve a problem closely related to WSD, namely, the lexical choice problem. We will examine these two different kinds of contextual information in this section.

6.2.1 Topical Context

With topical representation of context, the context of a given sense of a target word is a bag of words without any structure. Information in topical context is generally quite helpful for WSD. For instance, consider Examples (16) and (17) extracted from the Brown corpus, each containing an instance of the ambiguous word *bass*.

(16) … for scintillating flights of meaningless improvisations, and he has a quiet way of getting back and restating the melody after the improvising is over. In this he is sticking with tradition, however far removed from it he may seem to be. SHEARING TAKES OVER George Shearing took over with his well disciplined group, a sextet consisting of vibes, guitar, **bass**, drums, Shearing's piano and a bongo drummer. He met with enthusiastic audience approval, especially when he swung from jazz to Latin American things like the Mambo. Shearing, himself, seemed to me to be playing better piano than in his recent Newport appearances. A very casual, pleasant program- one of those easy-going things that make Newport's afternoon programs such a …

(17) … Breakfast was at the Palace Hotel, luncheon was somewhere in the mountain forest, and dinner was either at Boulder Creek or at Santa Cruz. Gazing too long at the scenery could be tiring, so halts were contrived between meals. Then the Chinese hostler, who rode with Vernon on the box, would break open a hamper and produce filets of smoked **bass** or sturgeon, sandwiches, pickled eggs, and a rum sangaree to be heated over a spirit lamp. In spring and in autumn the run was made for a group of botanists which included an old friend of mine. They gathered roots, bulbs, odd ferns, leaves, and bits of resin from the rare Santa Lucia fir, which exists only on a forty-five mile strip on the westerly side of these mountains. In the Spanish …

Intuitively, the first instance of *bass* can be disambiguated as INSTRUMENT-bass since *guitar*, *drum*, *piano*, *jazz*, etc. are likely to appear in the topical context of INSTRUMENT-bass. Similarly, the second instance can be disambiguated as FISH-bass since *meal*, *sandwiches*, *egg*, etc. are often found in the topical context of FISH-bass. Generally, the sense representation of topical context is acquired from a very large corpus. Gale, Church and Yarowsky [1992b] experimented on acquiring topical context from a substantial bilingual training corpus and reported good results.

6.2.2 Local Context

Local context includes structured information about word order, distance, and syntactic features. For instance, the local context of *a line from* does not suggest the same sense for the word *line* as *a line for* does.

Trigram as Local Context   Brown *et al.* [1990] used the trigram model as a way of resolving sense ambiguity for lexical selection in statistical machine translation. This model makes the assumption that only the previous two words have any effect on the translation, and thus, the word sense of the next word. The model was used to attack the problem of lexical ambiguity and produced satisfactory results, under some strong assumptions. For instance, the authors showed that the French sentence *Je vais prendre la decision* could be correctly translated as *I will make the decision* using this model. Although in isolation, *take* was more likely than *make* to translate as *prendre*, the trigram language reversed the decision in favor of *make*. A major problem with the trigram model is long distance dependency. For instance, the model incorrectly rendered the French sentence *Je vais prendre ma propre decision* as *I will take my own decision*. The language model did not consider *make my own decision* more probable since *prendre* and *decision* did not fall within a window of three words.

Lexical Relation   Dagan, Itai and Schwall [1991] and Dagan and Itai [1994] made use of translations of different senses from a Hebrew/English bilingual dictionary to disambiguate contexts. Local context in the form of lexical relations was analyzed in a foreign corpus. The basic idea of the algorithm is best explained with an example. Given two Hebrew words *hoze* and *shalom*, *hoze* has two translations in English: *contract* and *treaty*, while *shalom* is often translated into English as *peace*. Their experiment showed that all instances of *peace* appear before *treaty* and none before *contract* in the corpus of English language. Therefore, the authors concluded that this instance of *hoze* in the phrase *hoze shalom* was best translated as *treaty*. The authors experimented on lexical choice with 105 Hebrew words and 54 German words from news articles. The precision rates achieved ranged from 75% to 92% for coverage rates between 59% and 70%.

Approximating Lexical Relation   Brown *et al.* [1991] described a statistical algorithm for partitioning the senses of a word into two groups. The authors used mutual information to find a local contextual feature that most reliably indicated which of the senses of the French ambiguous word was used. For instance, for the verb *prendre*, the object was a good indicator: *prendre une measure* translated as *to take a measure,* and *prendre une decision* as *to make a decision*. Therefore, words (any word, first verb or first noun) immediately to the left or right of the word were evaluated for their effectiveness as good indicators for WSD and lexical choice. The authors reported 20% improvement in the performance of a machine translation system (from 37 to 45 sentences correct out of 100) when the words were first disambiguated in this way.

7 Conclusions

We have described an adaptive approach to word sense disambiguation. Under this new learning strategy, a contextual representation for each sense discriminator is first built based on the sense definition and example sentence in MRD and represented as a weighted-vector of concepts represented by word lists in a thesaurus. This knowledge representation acquired through MRD is based on a limited number of concepts; thus, the dreaded problem of data sparseness is avoided. Conceptual knowledge also offers the additional advantages of reduced storage requirements and increased efficiency due to reduced dimensionality. Also, we can correctly identify at least 50% of the word senses in unrestricted texts. In addition, these disambiguated texts can be used to adjust the fundamental knowledge in an adaptive fashion so to improve disambiguation precision. We have demonstrated that this approach can outperform established static approaches based on direct comparison of results obtained for the same words. This level of performance is achieved without lengthy training or the use of a very large training corpus.

Appendix A

A Glossary of LLOCE Topics

Here, we list 129 topics found in LLOCE. The column labeled "Topic" shows a set of two-character symbols representing the topics in LLOCE. Each topic is giving a gloss.

Topic	Glossary	Topic	Glossary
Aa	Life and living things	Bg	Bodily states and associated activities
Ab	Living creatures generally	Bh	Bodily conditions relating to health, sickness, and disability
Ac	Animals/Mammals	Bi	Diseases and ailments
Ad	Birds	Bj	Medicine and general medical care
Ae	Reptiles and amphibians	Ca	People
Af	Fish and other water creatures	Cb	Courting, sex, and marriage
Ag	Insects and similar creatures	Cc	Friendship and enmity
Ah	Parts of animals	Cd	Death and Burial
Ai	Kinds and parts of plants	Ce	Social organization in groups and places
Aj	Plants generally	Cf	Government
Ba	The body generally	Cg	Politics and elections
Bb	The body: overall	Ch	Political tension and trouble
Bc	The head and the face	Ci	Social classifications and situations
Bd	The trunk, arms, and legs	Cj	Law and order generally
Be	The skin, the complexion, and the hair	Ck	Courts of law and legal work
Bf	Fluids and waste products of the body	Cl	The police, security services, crime, and criminals

Topic	Glossary	Topic	Glossary
Cm	Prison and punishment	Gh	General grammatical words
Cn	Warfare, defence, and the army	Ha	Substances and materials generally
Co	The armed forces	Hb	Objects generally
Cp	Religion and beliefs	Hc	Specific substances and materials
Da	Architecture and kinds of houses and buildings	Hd	Equipment, machines, and instruments
Db	Parts of houses	He	Tools
Dc	Areas around and near houses	Hf	Containers
Dd	Residence	Hg	Electricity and electrical equipment
De	Belonging and owing, getting and giving	Hh	Weapons
Df	Furniture and household fittings	Ia	Making things
Dg	Clothes and personal belongings	Ib	Arts and crafts
Dh	Cleaning and personal care	Ic	Science and technology
Ea	Food generally	Id	Industry and work
Eb	Food	Ie	Education
Ec	Drinks	Ja	Numbers and quantities
Ed	Cigarettes and drugs	Jb	Mathematics
Ee	The preparation and quality of food	Jc	Measurement
Ef	Places and people associated with food and drink	Jd	Money
Eg	Farming	Je	Banking, wealth, and investment
Fa	Feeling and behavior generally	Jf	Commerce
Fb	Liking and not liking	Jg	Shopping and general expenses
Fc	Good and evil	Jh	Business, work, and employment
Fd	Happiness and sadness	Ka	Entertainment generally
Fe	Anger, violence, stress, calm, and quietness	Kb	Music and related activities
Fh	Kindness and unkindness	Kc	Recording sound, listening to the radio, etc.
Fi	Honesty, loyalty, trickery, and deceit	Kd	Drama, the theatre, and show business
Fj	Relaxation, excitement, interest, and surprise	Ke	Sport and games generally
Fk	Actions of the face related to feelings	Kf	Indoor games
Fl	Senses and sensations	Kg	Children's games and toys
Ga	Thinking, judging and remembering	Kh	Outdoor games
Gb	Knowing and learning	La	The universe
Gc	Communicating, mainly by speaking and talking	Lb	Light and color
Gd	Communicating, mainly by reading and writing, printing and publishing, radio and television	Lc	Weather and temperature
Ge	Communication and information	Ld	Geography
Gf	Language	Le	Time generally
Gg	Grammar	Lf	Beginning and ending

Topic	Glossary	Topic	Glossary
Lg	Old, new, and young	Nb	Possibility, chance, and necessity
Lh	Periods of time and their measurement	Nc	General, usual, unusual, etc.
Li	Grammatical words and phrases relating to time	Nd	Size, importance, and availability
Ma	Moving, coming, and going	Ne	Doing things
Mb	Putting and taking, pulling and pushing	Nf	Causing
Mc	Travel and visiting	Ng	Resemblance, difference, and change
Md	Vehicles and transport on land	Nh	Rightness, fairness, purpose, use, and strength
Me	Places	Ni	Fullness, heaviness, thickness, stiffness, roughness, etc.
Mf	Shipping	Nj	Actions and positions
Mg	Aircraft	Nk	Cutting, joining, breaking, and destroying
Mh	Location and direction	Nl	Showing, hiding, finding, saving, and similar words
Na	Being, becoming, and happening		

References

Benson, M., E. Benson and R. Ilson. *The BBI Combinatory Dictionary of English,* John Benjamins Publishing Company. Amstersam/Philadelphia. 1993.

Black, E. "An Experiment in Computational Discrimination of English Word Sense. " *IBM Journal of Research and Developmen.* Vol. 32. pp. 185-194. 1988.

Brill, E. *A Corpus-Based Approach to Language Learning,* PH. D. thesis. Department of Computer and Information Science. University of Pennsylvania. 1993.

Brown, P. F., J. Cocke, S. A. Della Pietra, V. J. Della Pietra, F. Jelinek, J. D. Lafferty, R. L. Mercer and P. S. Roosin. "A Statistical Approach to Machine Translation." *Computational Linguistics,*.16(2). pp. 79-85. 1990.

Brown, P. F., S. A. Della Pietra, V. J. Della Pietra and R. L. Mercer. "Word–Sense Disambiguation using Statistical Methods" In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistic.* pp. 264-270. 1991.

Chang, J. S., R. H. Shu and M. H. Chen. "Automatic Extraction Rules on Preposition Phrase." In *Proceedings of R.O.C. Computational Linguistics Conference IX* (*ROCLING-IX*). pp. 295-320. 1996.

Chen, J. N. and J. S. Chang, "A Concept-based Adaptive Approach to Word Sense Disambiguation. " In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics.* pp. 237-244. Montreal. Canada. 1998a.

Chen, J. N. and J. S. Chang. "TopSense: A Topical Sense Clustering Method based on Information Retrieval Techniques on Machine Readable Resources." *Special Issue on Word Sense Disambiguation. Computational Linguistics.* 24(1). pp. 61-95. 1998b.

Chen, J. N., J. S. Chang, H. H. Sheng and S. J. Ker. "Combining Machine Readable lexical Resources and Bilingual Corpora for Broad Word Sense Disambiguation." In *Proceedings of the Second*

Conference of the Association for Machine Translation. pp. 115-124. Montreal. Quebec. Canada. 1996.

Dagan, I. and A. Itai. "Word Sense Disambiguation Using a Second Language Monolingual Corpus." *Computational Linguistic.* 20(4). pp. 563-596. 1994.

Dagan, I., A. Itai, and U. Schwall. "Two Languages are More Informative than One." In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics.* pp. 130-137. 1991.

Gale, W. A., K. W. Church and D. Yarowsky. "One Sense Per Discourse," In *Proceedings of the Speech and Natural Language Workshop.* pp. 233-237. 1992a.

Gale, W. A., K. W. Church, and D. Yarowsky. "Using Bilingual Materials to Develop Word Sense Disambiguation Methods." In *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation.* pp. 101-112. 1992b.

Guthrie, J., L. Guthrie, Y. Wilks and H. Aidinejad. "Subject-dependent Co-occurrence and Word Sense Disambiguation." In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics.* pp. 146-152. 1991.

Hearst, M. "Noun Homonym Disambiguation using Local Context in Large Text Corpora." In *Proceedings of the 7th International Conference on of UW Centre for the New OED and Text Research: Using Corpora.* pp. 1-22. 1991.

Hawking, D. and P. Thistlewaite. "Proximity Operators – So Near and So Far," In *Proceedings of the fourth Text REtrieval Conference* (*TREC-4*). pp. 1-13. 1995.

Kelly, E. and P. Stone. *Computer Recognition of English Word Senses.* North-Holland. Amsterdam. 1975.

Leacock, C., G. Towell and E. M. Voorhees. "Corpus-based Statistical Sense Resolution." In *Proceedings of the ARPA Workshop on Human Language Technology.* 1993.

Lesk, M. E. "Automatic Sense Disambiguation using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone," In *Proceedings of the ACM SIGDOC Conference.* pp. 24-26, Toronto. Ontario. 1986.

Luk, A. K. "Statistical Sense Disambiguation with Relatively Small Corpora using Dictionary Definitions." In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics.* pp. 181-188. 1995.

McArthur, T. *Longman Lexicon of Contemporary English.* Longman Group (Far East) Ltd.. Hong Kong, 1992.

Proctor, P. (ed.) *Longman Dictionary of Contemporary English.* Harlow: Longman Group. 1978.

Proctor, P. (ed.) *Longman English-Chinese Dictionary of Contemporary English.* Longman Group (Far East) Ltd.. Hong Kong. 1988.

Roget's Thesaurus of English words and Phrases. Longman Group UK Limited. 1987.

Slator, B. "Using Context for Sense Preference." In Zernik (ed.) *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon.* Lawrence Erlbaum. Hillsdale. NJ. 1991.

Veronis, J. and N. Ide. "Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries." In *Proceedings of the 13th International Conference on Computational Linguistics.* pp. 389-394. 1990.

Walker, D. and R. Amsler. "The use of Machine-Readable Dictionaries in Sublanguage Analysis." In *Analyzing Language in Restricted Domains*. Grishman. R. and R. Kittredge (eds.). Lawrence Erlbaum Associates. Hillsdale. New Jersey. 1986. (also available in R. Kittredge (ed.), *Proceedings of Workshop on Sublanguage Analysis*; New York 1984).

White, J. S. "Determination of Lexical-Semantic Relations for Multi-Lingual Terminology Structures." In *Relational Models of the Lexicon*. Cambridge University Press, Cambridge. UK. 1988.

Wilks, Y. A., D. C. Fass, C. M. Guo, J. E. McDonald, T. Plate and B. M. Slator. "Providing Tractable Dictionaty Tools." *Machine Translation*. 5. pp. 99-154. 1990.

Yarowsky, D. "Word-Sense Disambiguation using Statistical Models of Roget's Categories Trained on Large Corpora." In *Proceedings of the 14th International Conference on Computational Linguistics*. pp. 454-460. Nantes. France. 1992.

Yarowsky, D. "Unsupervised Word Sense Disambiguation Rivaling Supervised Methods." In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*. pp. 189-196. 1995.

Zipf, G. *Human Behavior and the Principle of Least Eeffect*. Hafner. New York. 1994.