WORD SENSE AMBIGUATION: CLUSTERING RELATED SENSES [1]

William B. Dolan

Microsoft Research
billdol@microsoft.com

Abstract

This paper describes a heuristic approach to automatically identifying which senses of a machine-readable dictionary (MRD) headword are semantically related versus those which correspond to fundamentally different senses of the word. The inclusion of this information in a lexical database profoundly alters the nature of sense disambiguation: the appropriate "sense" of a polysemous word may now correspond to some *set* of related senses. Our technique offers benefits both for on-line semantic processing and for the challenging task of mapping word senses across multiple MRDs in creating a merged lexical database.

1. Introduction

The problem of word sense disambiguation is one which has received increased attention in recent work on Natural Language Processing (NLP) and Information Retrieval (IR). Given an occurrence of a polysemous word in running text, the task as it is generally formulated involves examining a set of senses, defined by a MRD or hand-constructed lexicon, and examining contextual cues to discover which of these is the intended one. This paper considers a problem with the standard approach to handling polysemy, arguing that in many cases this kind of "forced-choice" approach to disambiguation leads to arbitrary decisions which have negative consequences for NLP systems. In particular, we show that a great deal of potentially useful information about a word's meaning may be missed if the task involves isolating a single "correct" sense. We describe an approach to the construction of an MRD-derived lexical database that helps overcome some of these difficulties.

We begin by reviewing two difficulties with this approach, then go on to suggest our approach to solving these difficulties in creating a large MRD-derived lexical database. Our method might be termed "ambiguation", because it involves blurring the boundaries between closely related word senses.

After describing the algorithm which accomplishes this task, we go on to briefly discuss its results. Finally, we describe the implications of this work has for the task of merging multiple.

The arbitrariness of sense divisions

The division of word meanings into distinct dictionary senses and entries is frequently arbitrary (Atkins and Levin, 1988; Atkins, 1991), as a comparison of any two dictionaries quickly makes clear. For example, consider the verb "mo(u)lt", whose single sense in the American Heritage Dictionary, Third Edition (AHD3) corresponds to two senses in Longman's Dictionary of Contemporary English (LDOCE):[2]

AHD3

> (1) "to shed periodically part or all of a coat or... covering, such as feathers, cuticle, or skin..."

LDOCE

> (1) "(of a bird) to lose or throw off (feathers) at the season when new feathers grow"
> (2) "(of an animal, esp. a dog or cat) to lose or throw off (hair or fur)"

The arbitrary nature of such divisions is compounded by the fact that dictionaries typically provide no information about how the different senses of a polysemous headword might be related. Examination of dictionary entries shows that these interrelationships are often highly complex, encompassing senses which differ only in some slight shade of meaning, those which are historically but not synchronically related, those which are linked through some more or less opaque process of metaphor or metonymy, and finally, those which appear to be completely unrelated.

A typical case is the entry for the noun "crank", which includes one "apparatus" sense (1) and two "person" senses (2) and (3). Nothing in this entry indicates that (2) and (3) are more closely related to one another than either is to (1).

---

[2] All examples are from LDOCE, except as noted.

(1) "an apparatus for changing movement in a
   straight line into circular movement..."
(2) "a person with...strange, odd, or peculiar ideas"
(3) "a nasty bad-tempered person"

Using MRDs for Sense Disambiguation

Atkins (1991) argues that dictionary-derived lexical databases will be capable of supporting high-quality NLP only if they contain highly detailed taxonomic descriptions of the interrelationships among word senses. These relationships are often systematic (see Atkins, 1991), and it is possible to imagine strategies for automatically or at least semi-automatically identifying them. One such proposal is due to Chodorow (1990), who notes 10 recurring types of inter-sense relationships in Webster's 7th, including PROCESS/RESULT, FOOD/PLANT, and CONTAINER/VOLUME, and suggests that some instances of these relationships might be automatically identified. Ideally, such strategy might allow the automated construction of lexical databases which explicitly characterize *how* individual senses of a headword are related, with these interrelationships described by a fixed, general set of semantic associations which hold between words throughout the lexicon.

In practice, however, attempts to automatically identify systematic polysemy in MRDs will capture only a small subset of the cases in which word senses overlap semantically. Often, distinctions among a word's senses are so fine or so idiosyncratic that they simply cannot be characterized in a general way. For instance, while the two LDOCE senses of "moult" are closely related, the fine distinction they reflect between "bird" and "animal" behavior is not one which recurs systematically throughout the English lexicon.

In short, the task of identifying and attaching a meaningful label to each of the links among related words senses in a large lexical database is a daunting one, and one that will ultimately require a great deal of hand-coding. Perhaps for these reasons, we know of no large-scale attempts to automatically create labeled links among senses of polysemous words.

Moreover, it is not clear that attaching a meaningful label to the relationship between two semantically related senses of a word will necessarily aid in performing NLP tasks. Krovetz and Croft (1992) suggest just the opposite, claiming that in many cases, dictionary entries for polysemous words encode fine-grain semantic distinctions that are unlikely to be of practical value for specific applications. Our experience suggests a similar conclusion. Consider, for instance, the following pair of senses for the noun "stalk":

(1) "the main upright part of a plant (not a tree)"
   (Ex: "a beanstalk")
(2) "a long narrow part of a plant supporting one or
   more leaves, fruits, or flowers; stem"

The differences between these two senses are subtle enough that for many tasks, including sense disambiguation in running text, the two are likely to be indistinguishable from one another. In a sentence like "The stalks remained in the farmer's field long after summer", for instance, the choice of some particular sense of "stalk" as the "correct" one will be essentially arbitrary.

Sense Disambiguation versus Information Loss

Sense disambiguation algorithms are frequently faced with multiple "correct" choices, a situation which increases their odds of choosing a reasonable sense, but which also has hidden negative consequences for semantic processing. First of all, the task of discriminating between two or more extremely similar senses can waste processing resources while providing no obvious benefit. However, there are more problematic effects of combining a lexicon which makes unnecessarily fine distinctions between word senses with a disambiguation algorithm which sets up the artificial task of choosing a single "correct" sense for a word. The problem is that this strategy means that the amount of semantic information retrieved for a word will always be limited to just that which is available in some individual sense, and valuable background information about a word's meaning may be ignored. In the case of "stalk", for instance, choosing the first sense will mean losing the fact that "stalks" are "stems", that they are "narrow", and that they "support leaves, fruits, or flowers". Choosing the second sense, on the other hand, will mean losing the fact that stalks are upright, that one example of a stalk is a "beanstalk", and that the main upright part of a "tree" cannot be called a "stalk".

Human dictionary users never encounter this problem. The reason is that instead of treating the entry for a word like "stalk" as a pair of entirely discrete senses, a human looking this word up would typically arrive at more abstract notion of its meaning, one which encompasses information from *both* senses. How can we reformulate the problem of sense disambiguation in a computational context so that semantic processing can do a better job of mimicking the human user? Our solution involves encoding in our LDOCE-derived lexical database information about how a word's senses overlap semantically.

2. Identifying Semantically Similar Senses

The remainder of the paper describes a heuristic-based algorithm which automatically determines which senses of a given LDOCE headword are

closely related to one another vs. those which appear to represent fundamentally different senses of the word. While no attempt is made to explicitly identify the *nature* of these links, our program has the advantage of generality: no hand-coding is required, and the techniques we describe can thus be applied (with some modification) to on-line dictionaries other than LDOCE. This work has an important effect on the formulation of the sense disambiguation task: by encoding information of this kind in our LDOCE-derived lexical database, we can now permit the sense disambiguation component of our system to return a merged representation of the semantic information contained in *multiple* senses of a word like "stalk". Making available more background information about a word's meaning increases the likelihood of correctly interpreting sentences which contain this word.

Our method involves performing an exhaustive set of pairwise comparisons of the different senses of a polysemous word with one another, with the aim of discovering which pairs show a higher degree of semantic similarity. Comparisons are not limited by part of speech; for example, noun and verb senses are compared to one another. A variety of types of information about a sense's meaning are exploited by this comparison step, including:

• LDOCE Syntactic Subcategorization Codes

• LDOCE Boxcodes

The program uses a taxonomic classification of these codes based on Bruce and Guthrie (1992) to allow partial matches between senses with non-identical but related Boxcodes. In addition, certain Boxcode specifications (e.g., [plant]) match against sets of keywords in definition strings (e.g., {*plant, soil*}).

• LDOCE Domain Codes

A taxonomic classification of the 124 Domain codes like that in Slator (1988) is used to identify cases in which two senses have similar but non-identical codes. As with the Boxcodes, certain Domain specifications (e.g., *BB*, "baseball"]) match against sets of keywords in definition strings (e.g., {*baseball, ball,sports*}).

• Features Abstracted from LDOCE Definitions:

A number of binary features, including [locative] and [human] have been automatically assigned to LDOCE senses, based on syntactic and lexical properties of their definitions. Matches between these features increase the likelihood that two senses are semantically related.

• Semantic Relations

The most important source of evidence about the interrelationships among senses has been automatically derived from LDOCE definition sentences. The program consults a lexical database

which contains approximately 150,000 semantic associations between word senses, the result of automatically parsing the definition text of each noun and verb sense in LDOCE and then applying a set of heuristic rules which automatically attempt to identify any systematic semantic relationships holding between a headword and the (base forms of) words used to define it (Jensen & Binot, 1987; Montemagni and Vanderwende, 1992). Approximately 25 types of semantic relations are currently identified, including *Hypernym* (genus term), *Location, Manner, Purpose, Has_Part, Typical_Subject*, and *Possessor*. Finally, each of these links is automatically sense-disambiguated. The resulting associations are modeled as labeled edges in a directed cyclic graph whose nodes correspond to individual word senses (Dolan et al, 1993; Pentheroudakis and Vanderwende, 1993).

Matching two senses involves comparing any values which have been identified for each of the semantic relation types. One of the most important comparisons is of *Hypernyms*, which have been identified for the vast majority of noun and verb senses. An exact *Hypernym* match generally signals a close semantic relationship between two senses, as in the following senses of the noun "cat":

(1) "a small *animal* with soft fur and sharp teeth and claws (nails), often kept as a pet..."
(2) "any of various types of *animals* related to this, such as the lion or tiger"

Comparisons are not limited to *Hypernyms*, of course: in comparing two senses, the program attempts to identify shared values for each of the different semantic attributes present in a word's lexical representation. For instance, in each of the following verb senses of "crawl", the word "slowly" has been automatically identified as the value of a *Manner* attribute.

(1) "to move *slowly* with the body close to the ground or floor, or on the hands and knees"
(2) "to go very *slowly*"

Each time an identical value is found for a given semantic attribute, the algorithm increments the correlation score for that pair of senses. If no exact match is found, the program checks whether the values for this attribute in the two senses have a hypernym or hyponym in common. The following senses of the noun "insect", for example, are linked through the *Hypernyms* "creature" and "animal":

(1) "a small *creature* with no bones and a hard outer covering..."
(2) "a very small *animal* that creeps along the ground, such as a spider or worm"

According to the network implicit in LDOCE, "creature" is a hyponym of "animal", while "animal" is a hyponym of "creature". (For discussion of this

kind of circularity in dictionary definitions, see Calzolari, 1977.)

In addition to such straightforward comparisons, a number of "scrambled" comparisons are attempted. For instance, any value for the *IngredientOf* attribute is automatically compared to the *Hypernym* value(s) for each other senses. This comparison reflects the fact that many nouns are both the name for a substance and for something which is made from that substance. An example of this is the noun "coffee": in one sense, "coffee" is *IngredientOf* of a "drink", while in another sense it has been identified as a *Hypernym* of the noun "drink".

(1) "a brown powder made by crushing coffee beans, *used for making drinks...*"
(2) "(a cupful of) a hot brown *drink* made by adding hot water and/or milk to this powder"

3. Discussion and Evaluation

The sense clustering program was run over the set of 33,000 single word noun defintions and 12,000 single word verb definitions in LDOCE (45,000 total) in a process that took approximately 20 hours on a 486/50 PC. Given a set of senses for a polysemous word such as "crank", the result of the exhaustive pairwise comparisons performed by the program is a (symmetrical) matrix of correlation scores:

	v (1a)	v(1b)	n 1	n 2	n 3
v (1a)					
v (1b)	75				
n 1	41	35			
n 2	3	3	2		
n 3	3	3	2	48	

Since our comparison are heuristic in nature, the relative rankings of the pairwise comparisons for a polysemous word's senses are the relevant measure of semantic similarity, rather than any absolute threshold. In the case of "crank", Clustering has correctly indicated a high correlation between the two "human" senses of the noun "crank", and a high correlation between the two verbal subsenses and the "apparatus" noun sense. Moreover, the two "human" noun senses are not semantically correlated with any of the three "apparatus" senses.

Negative scores are also common, reflecting certain kinds of incompatibilities between senses (e.g., one sense is [+*animate*] while the other is [-*animate*]). As a rule of thumb, however, it is much easier to identify commonalties between senses than to identify definite mismatches.

One of the most useful products of clustering is the identification of many cases of zero-derived noun/verb pairs. For instance, the comparison of the various senses of the word "cook" shows the verb sense "to prepare (food) for eating..." to be highly correlated with the noun sense "a person who prepares and cooks food". This kind of cross-classification, which dictionaries generally fail to provide, has interesting implications for normalizing the semantics of superficially very different sentences. For example, a concept which is expressed verbally in one sentence can now be related to the same general concept expressed nominally in another, even if LDOCE does not explicitly link the definitions for the two parts of speech. (Pentheroudakis & Vanderwende (1993) describe a general approach to identifying semantic links among morphologically-related words.)

Metaphor

Interestingly, the fact that many conventional metaphors are lexicalized in dictionary definitions can lead to difficulties with our strategy of comparing different definitions to one another.[3] Consider the following senses of the noun "mouth":

(1) "the opening on the face through which an animal or human being may take food..."
(2) "an opening, entrance, or way out"
   (*Ex: "mouth of a cave"*)

In considering these two senses, Clustering returned a correlation score of 26, suggesting a reasonably close semantic relationship between them. From one perspective this is simply wrong: a human or animal "mouth" is fundamentally different from a cave "mouth", and we would like our MRD-derived lexicon to indicate this fact. Once the obvious metaphorical association between these two senses of "mouth" is noted, however, the reason for the clustering program's result becomes clear: both senses are defined as kinds of "openings". The case for treating the two senses as semantically similar is strengthened by other evidence: one sense of "entrance" (which is the *Hypernym* of the second sense) has "opening" as its own *Hypernym*: "a gate, door, or other *opening* by which one enters".

Such metaphorical associations between word senses add a considerable degree of complexity to disambiguation and other kinds of reasoning processes that operate by identifying semantic relationships between different words. More work aimed at identifying the systematic nature of such

---

[3] This same problem crops up in any task which involves comparing different words/senses, including disambiguation of running text.

relationships will be required before metaphor-based confusions of the kind described above can be automatically resolved.

4. Conclusions and Future Work

Interestingly, the machinery used to identify common semantic threads among a polysemous word's senses was originally constructed with another purpose in mind -- namely, disambiguating LDOCE genus terms. As it turned out, exactly the same set of tests used to compare a word sense to the set of possible senses of its *Hypernym* proved useful in comparing the different senses of a single word. While the current instantiation of the Clustering program relies partially on information which is idiosyncratic to LDOCE (e.g., Domain codes), most of the information it uses for inter-sense comparisons has been extracted from the text of their definitions. For this reason, the techniques we have described here are can be readily applied to other MRDs.

In addition, we plan to experiment with augmenting the results of our sense clustering with statistics derived from running a sense disambiguation program over a large free-text corpus. In particular, we are interested in discovering whether the "hard cases" encountered by this sense disambiguation program (i.e., those cases in which the program consistently has difficulty in choosing among two or more competing senses) correlate with cases of significant semantic overlap among senses. If this hypothesis is borne out, then information about which senses are difficult to distinguish in free text can be used to help us establish the taxonomic relationships among the different senses of a polysemous word.

Finally, the work we have described here has important implications for the task of merging multiple MRDs into a single lexical database. This task is greatly complicated by arbitrary sense divisions encountered in different dictionaries (see Atkins and Levin, 1988; Byrd, 1989). Consider the verb "mo(u)lt" again: since the single AHD3 sense for this word subsumes both LDOCE senses, no obvious mapping strategy is available. Should the AHD3 sense be mapped into just one of the LDOCE senses? Each of them? Or should the AHD3 sense be left separate, resulting in a merged lexical entry with three separate entries? As more sources of information about word meanings are folded in, this last strategy can only increase the complexity of semantic processing, since it will become more and more difficult to determine which of an ever-larger set of semantically-related senses is the appropriate one in a given context. Clustering offers a simple way to begin to approach this problem. By pooling and clustering senses for words from both LDOCE and AHD3, we can provide a rough indication of the semantic interconnections between the two entries.

As our techniques for automatically extracting semantic information from the text of definition and example sentences gradually improve, we expect our ability to automatically identify semantic overlaps and differences to improve as well.

5. References

Atkins, Beryl (1991) "Building a Lexicon: the contribution of lexicography", *International Journal of Lexicography* 3: 167-204.

Atkins, B. and B. Levin (1988) "Admitting Impediments", *Proceedings of the Fourth Annual Conference of the UW Centre for the New OED*, Oxford.

Bruce, R. and L. Guthrie (1992) "Genus Disambiguation: a Study in Weighted Preference", In *Proceedings of COLING92*, pp. 1187-1191.

Byrd, R. (1989) "Discovering Relationships among Word Senses", *Proceedings of the Fifth Annual Conference of the UW Centre for the New OED*, Waterloo, Canada.

Calzolari, N. (1977) An Empirical Approach to Circularity in Dictionary Definitions", in *Cahiers de Lexicologie* 31: 118-128.

Chodorow, M. (1990) "Making Sense of Word Senses: detecting and analyzing systematic polysemy in noun definitions", CUNY, unpublished ms.

Dolan, W., L. Vanderwende, and S. Richardson (1993). "Automatically Derived Structured Knowledge Bases from On-line Dictionaries", *Proceedings of the First Conference of the Pacific Association for Computational Linguistics*, April 21-24, Simon Fraser University, Vancouver, Canada.

Jensen, K. and J.-L. Binot (1987). "Disambiguating prepositional phrase attachments by using on-line dictionary definitions". *Computational Linguistics* 13: 3-4, pp. 251-260.

Krovetz, R. and B. Croft (1992). "Lexical Ambiguity and Information Retrieval", *ACM Transactions on Information Systems*, 10: 2, pp. 115-141.

Montemagne, S. and L. Vanderwende (1992). "Structural patterns vs. string patterns for extracting semantic information from dictionaries." In *Proceedings of COLING92*, pp. 546-552.

Pentheroudakis, J. and L. Vanderwende (1993) "Automatically Identifying Morphological Relations in Machine Readable Dictionaries", in *Proceedings of the Ninth Annual Conf. of the UW Centre for the New OED and Text Research*, Oxford, England.pp. 114-131.

Slator, B. (1988) "Constructing Contextually Organized Lexical Semantic Knowledge-Bases", *Proceedings of the Third Annual Rocky Mountain Conference on Artificial Intelligence* (RMCAI-88), Denver, CO, pp. 142-148.