Natural Language Processing, Disambiguation in

Intermediate article

Graeme Hirst, University of Toronto, Toronto, Ontario, Canada

CONTENTS

Ambiguity pervades all levels of language. Any computer system that uses the meaning of natural language must disambiguate its input.

INTRODUCTION

Language is rife with ambiguity: a single utterance can, in principle, have many different interpretations or meanings. Usually, however, the speaker or writer intends just one of these meanings – usually, only one of them will make sense – and humans are adept at rapidly determining which one was intended, not even consciously noticing the others. Computer language-processing systems that use the meaning of an utterance in their task must therefore, like people, disambiguate their input. (*See* **Natural Language Processing**)

For example, an English-to-French machine translation system must decide whether to translate the word *duty* as *douane* or *devoir*, depending on whether its meaning in the utterance relates to an import tax or to a moral or legal obligation. A speech recognition system that hears [djuti] must decide whether it should be rendered as *duty* or *due tea*. Suppose a text interpretation system comes across this excerpt (from Jane Austen's *Persuasion*):

> The Admiral, after taking two or three refreshing turns about the room with his hands behind him, …

It must then decide whether the prepositional phrase *with his hands behind him* describes the room or the Admiral's manner of taking refreshing turns. And in either case, it must also decide exactly whose hands are behind whom. (*See* **Machine Translation**)

As these examples show, ambiguity can occur in many ways and at all levels of language. This article will cover ambiguities of syntax and of word senses, including the special case of pronouns. Most of the article applies to both spoken and written text, and the terms *speaker* and *writer* will be used interchangeably. Ambiguity can be treated more specifically at the phonetic level and at the morphological level. (*See* **Speech Perception and Recognition, Theories and Models of**; **Morphological Processing**)

LEXICAL DISAMBIGUATION

There are two kinds of lexical ambiguity – that is, of ambiguity of words. The first is ambiguity as to the syntactic category, or part of speech, of a word in an utterance: for example, the word *flies* can be used as either a verb or a noun. While only about 12 percent of the word types of English are ambiguous as to category, they tend to be the more common words, representing about 40 percent of the word tokens that are uttered (DeRose, 1988). The second kind of lexical ambiguity is ambiguity of meaning even after the part of speech is determined. (*See* **Lexical Ambiguity Resolution**)

Part-of-speech Tagging

Ambiguities of syntactic category are resolved as part of the process of 'part-of-speech tagging' – labeling each word in an input sentence with its category – which is the first stage of processing in many applications of natural language processing. The rules of grammar constrain the allowable sequences of syntactic categories – in English, for example, the base form of a verb may not immediately follow the definite article *the* – and about 60 percent of word tokens are not ambiguous as to category (including, in English, the articles *the* and *a*). Consequently, the category of a word can be resolved with a high degree of accuracy, 97 percent or better, just by looking at the categories of a few preceding words. It is easy to construct examples in which a much greater number is required, but such cases are rare in practice. Part-of-speech tagging is usually regarded as a probabilistic process in which the most likely category is chosen in light of the two preceding words and the potential categories of the word under consideration. Two sources of information are thus required: a lexicon of words, listing the allowable categories of each, and knowledge of allowable sequences of categories along with their probabilities of occurrence. (*See* **Natural Language Processing, Statistical Approaches to**; **Lexicon**; **Lexicon, Computational Models of**)

The Nature of Word Senses

A glance at a page of a dictionary reminds us that it is only a small minority of words – mostly technical terms – that have only a single sense. When the senses of a word are closely related, the word is said to be polysemous. For example, *window* can mean either an opening in a wall (*crawled through the window*) or the glass that fits in the opening (*broke the window*). When the senses are completely different, the word is said to be homonymous. For example, *ash* can mean either a tree or the residue from combustion. A single word may be both homonymous and polysemous: *bank* is homonymous in its senses relating to financial institutions and to the edge of a watercourse; but, when pertaining to a financial institution, it is polysemous in that it can denote both the institution and the building in which the institution does business. In speech recognition, word sense ambiguity arises from similarity of sound rather than spelling; thus disambiguation is required between *see* and *sea*.

The conventional view of word senses assumes that for each word there is a fixed inventory of senses to decide among, and that for any particular utterance, to disambiguate is to choose exactly one of these senses as 'correct'. The assumption of a fixed inventory of senses has been challenged by many researchers (e.g. Kilgarriff, 1997), who point to the wide disparities in treatments of the same word by different lexicographers in different dictionaries, especially with regard to fine-grained distinctions between polysemous senses: what is one sense for one lexicographer might be two or three for another. And people often find it hard to decide which fine-grained dictionary sense best represents the meaning of a word in its context (Kilgarriff, 1992); often, they will say that a word is being used in two senses at once. For example, *Nadia visited the bank to get some money* seems to invoke *bank* as both building and institution simultaneously. Can we reasonably expect or require a computer to be more precise or decisive about word senses than people are? Often, it doesn't matter: in many applications of natural language processing, very fine-grained word sense disambiguation is unnecessary, and it suffices to resolve homonymy and perhaps coarse-grained polysemy at a level where there is reasonable agreement as to the inventory of senses. For example, a program that translates English to French needs to know whether an occurrence of *bank* is used in a financial or river-related sense in order to choose the correct translation; but, if it is used in a financial sense, the program need not decide between the institution and the building as the translation is the same in either case. (*See* **Word Meaning, Psychology of**; **Word Recognition**)

Methods of Word Sense Disambiguation

When an ambiguous word is a member of more than one syntactic category, part-of-speech tagging allows senses not associated with the category in which the word is being used to be eliminated from consideration; occasionally this yields a unique sense. A unique sense can also be assigned if the word is recognized as part of an unambiguous lexicalized compound; for example, if *private school* is listed as a phrase with its own meaning, there is no need to choose among the different senses of *private* and of *school*. But usually more sophisticated methods are required.

Many such methods are based on selectional restriction, relationship to the topic of the text, or both. In addition, the relative frequency of two different senses may be used to break a tie between them when other methods are unable to choose. In particular, when one or more of a word's senses are relatively rare, they may be eliminated from consideration unless there is positive evidence for them: for example, the noun *email* in its now-rare sense of enamel should be ruled out in favor of the electronic mail sense unless there is some particular reason to suppose that enamel is intended. Data on how frequent each sense of a word is can be derived from large corpora of text that have been disambiguated by humans or by semi-automatic methods with human verification. The accumulation of sufficient sense frequency data to adequately cover a language is an enormous task, however, and so far only relatively small sense-tagged corpora exist for English (Resnik and Yarowsky, 1999). Like relative frequency, the 'one sense per discourse' heuristic (Yarowsky, 1995) can serve as an adjunct to any other method. This heuristic relies on the fact that it is rare in practice for a homonym to be used in more than one sense within the same text or discourse; so if, for example, the word *crane* occurs five times in a text and some disambiguation method deems it to be a bird in four instances and construction equipment in the other, then the latter is almost certainly wrong and should be corrected to the majority vote.

Selectional restrictions are the semantic constraints that a word sense may place on the senses of other words that combine with it. For example, the verb *eat* requires in literal language that its subject be an animate being and its object be something edible; so in *the mouse ate the corn*, we favor *mouse* as rodent rather than computer equipment and *corn* as cereal rather than callus. Metaphor and other kinds of nonliteral language can violate selectional restrictions (*the photocopier ate my report*), so such restrictions are helpful but not absolute constraints. Selectional restrictions vary in their degree of specificity: *elapse* accepts only time or a unit of time as its subject, whereas many different kinds of things can *grow*.

For a natural language processing system to use selectional restrictions, it first needs a knowledge base of the restrictions pertaining to each word sense, but no such knowledge base yet exists, and the creation of such a resource would be a large and poorly defined lexicographical task (the FrameNet project (Johnson and Fillmore, 2000) is a step in this direction). Resnik (1998) has proposed a process that can construct such a knowledge base automatically from a parsed corpus and an online hierarchical thesaurus such as WordNet (Fellbaum, 1998). For example, if the corpus contains examples of forms of the verb *drink* with objects such as *coffee*, *wine*, and *water*, the process can learn, by looking up these words in the thesaurus, that *drink* tends to select objects that are beverages or liquids. Resnik's experiments with the process showed that the information it derived was helpful, but of course not by itself sufficient for reliable disambiguation of word senses.

Many methods of word sense disambiguation have tried to capture the intuition that a good disambiguation cue, especially for homonyms, is the existence of a general semantic relationship between one of the candidate senses and those of nearby words in the text. For example, in proximity to the words *garden* and *pest*, the word *mole* is much more likely to refer to a mammal than to a skin blemish or a chocolate sauce. More generally, the topic of the text as a whole can be a helpful cue. The problem is how to make this idea precise and determine the semantic relationships.

Lesk (1986) proposed that dictionary definitions could be used for this. From an online dictionary, the definitions of all content words within, say, 100 words of the target word are found. Regarding this set of definitions as nothing more than a 'bag of words', with no consideration of the structure of the sentences or even the order of the words, the candidate sense of the target word is chosen that contains in its own definition more of the words in the bag than any of the other candidates. For a simplified example with just one word of context, consider the word *keyboard* in the phrase *the keyboard of the terminal*: its dictionary definition includes, in one of its senses, the word *computer*, as does one of the senses of *terminal*; accordingly, this sense of *keyboard* is chosen. Observe that *terminal* is similarly disambiguated. Lesk's method is surprisingly effective given its simplicity, and serves as the baseline against which more complex methods are compared (Kilgarriff and Palmer, 2000).

An example of a more complex method is the use of naive Bayesian classification to classify words according to which sense of each ambiguous word they tend to be associated with. For example, *money* tends to be associated with the financial sense of *bank*, and so do the words *loan* and *mortgage*, but *time* does not and *grass* is probably a contraindication. By looking at a very large corpus of text in which each word is tagged with its correct sense, and counting the number of times that each sense occurs with various other words in its proximity, we can compute the probability of any given word occurring in the proximity of each sense. Then, when disambiguation is necessary, the probability of each sense can be computed in the context of the nearby words, even if those words do not all indicate the same sense, and the sense with the greatest probability can be chosen. This method assumes that all the words in the context are conditionally independent of one another: the probability of seeing one word in context is independent of seeing any other word in the same context. Obviously, this is not true in practice, almost for the very reason that we want to use this method: words of related meaning tend to cluster. None the less, the method gives reasonable results. (*See* **Machine Learning**; **Natural Language Processing, Statistical Approaches to**)

However, this method is limited by the need for sense-tagged corpora as training data. Sufficient data do not exist to cover English, let alone less studied languages. Researchers have sought methods of circumventing this limitation. Yarowsky (1992) proposed that naive Bayesian classification could be used if the goal is not to determine the fine-grained sense of an ambiguous word but merely an indication of the topic with which it is associated: in effect, resolution of homonyms, which, while coarse-grained, is none the less useful in many applications such as information retrieval. For example, instead of having to determine separately the probability that the word *money* indicates a certain sense of *bank* and so do *deposit* and *account* and *river* and *canal* and *creek*, we instead determine that any word related to finance indicates one sense of *bank* (or one group of senses) and any word related to watercourses indicates another. Yarowsky used the categories of *Roget's Thesaurus* as his set of topics. In an experiment on 12 ambiguous words that appeared in a total of 39 thesaurus categories, he determined, from a corpus of 10 million words, what other words were both frequent and salient as indicators of each of the thesaurus categories in which those words appeared; he then used these words in a naive Bayesian process to classify occurrences of the same 12 words in a test corpus. The results were very good for words such as *mole*, whose senses are generally topic-specific, but not for words such as *interest*, whose senses tend to cut across topics.

But while this method avoids the need for a sense-tagged corpus, it still requires supervised training – that is, its learning phase is still based on some predefined knowledge source, in this case the thesaurus. Yarowsky (1995) has also proposed a method by which decision lists for disambiguation can be learned by unsupervised training. A decision list is an ordered sequence of very specific conditions for classifying a word by meaning: for example, a decision list for the word *bass* might include the conditions 'if the next word is *player*, the topic is music' and 'if the next word is *are*, the topic is fish'. The list is derived from an extremely large corpus, along with a 'seed' – an extremely strong cue – for each sense of the ambiguous word (*bird* and *construction* could be seeds for *crane*). Because the corpus is so large – 460 million words in Yarowsky's experiments – the seeds are sufficient to indicate a number of definite occurrences of each sense, whose context words, in turn, suggest additional cues to each sense. When some of the data have been thus tagged, a classification algorithm is used to find additional rules. The process then iterates, alternating with the 'one sense per discourse' heuristic, until most or all occurrences of the ambiguous word in the corpus have been tagged. The resulting decision lists give a disambiguation accuracy similar to that of the thesaurus-based method.

Both of Yarowsky's methods require separate training for each ambiguous word, so in practice they have been tried only on a few test words. The task of using these methods to cover all ambiguous words of a language remains a daunting one.

STRUCTURAL DISAMBIGUATION

Structural ambiguity is ambiguity of the structure of the utterance itself, as seen in the 'Admiral' example above, in which the prepositional phrase *with his hands behind his back* could be a modifier of *taking*, describing the manner in which the turns around the room were taken, or of *room*, describing the room. The ambiguity in this example is often referred to as one of 'prepositional phrase attachment', as the problem is determining which node in the parse tree of the sentence the prepositional phrase should be attached to. There are many kinds of structural ambiguity (the attachment point of relative clauses is another important one) but prepositional phrase attachment in English has received the most study and will be the example used here. (*See* **Sentence Processing**; **Sentence Processing: Mechanisms**)

Because the ambiguity is reflected in the parse tree of the sentence, its resolution is part of the process of syntactic analysis of the sentence, or parsing. A common way to conceive of the problem is that the parser determines, from the rules of syntax of the language, what the possible attachment points are, and then asks some other process to determine which is most likely to be correct in context (Hirst, 1987). (However, in the case of lexically conditioned statistical parsers, such as that of Collins (1996), no distinction is made between attachment decisions that are mandated by the grammar of the language and those, the kind that are of interest here, that are 'discretionary'; in both cases, the decisions are based on the probabilities of syntactic dependencies between the particular words.) Structural ambiguity and lexical ambiguity are independent phenomena, but clearly resolution of either one interacts with resolution of the other: the best attachment point might depend on the meaning of a word, and the most likely meaning of a word might depend on a structural decision. In practice, however, the two ambiguities are usually considered separately. (*See* **Parsing**)

Prepositional phrase attachment ambiguity, in its simplest form, has three or four elements: a verb (e.g. *await*), the head noun of its object (e.g. *approval*), the preposition (e.g. *from*), and, in some methods, the head noun of the prepositional phrase (e.g. *government*). The disambiguation process must choose between the verb and the object head noun as the attachment point. Recent approaches have tried, in various ways, to use the relative frequency of each attachment, as determined by statistics gathered from a large corpus of sentences. Because manual annotation of the corpus is not required, more data are available for this than for the analogous problem in lexical disambiguation. However, the problem is harder because even in very large corpora, most combinations of three elements occur rarely if at all; using four elements instead of three increases the potential accuracy of the method at the expense of exacerbating the sparseness of the data.

Taking the three-element problem, Hindle and Rooth (1993) achieved about 80 percent accuracy with an unsupervised training method based on the attachment probabilities that were observed in a 13-million-word corpus of newswire text. The corpus had been almost fully parsed but lacked resolution of its ambiguous prepositional phrase attachment points. The method computed 'lexical association (LA) scores', defined as the logarithm (base 2) of the relative likelihood of verb and noun attachment for triples: for example, LA(*send*, *soldier*, *into*) was found to be approximately 5.81, meaning that verb attachment is 56 (i.e. $2^{5.81}$) times more likely than noun attachment in the sentence *Moscow sent more soldiers into Afghanistan*. These scores were determined by first gathering data from cases of unambiguous prepositional phrase attachment in the corpus (such as attachments to subjects of sentences and attachments to verbs without objects) and then, where strong lexical associations were found, using these data to resolve ambiguous cases; the procedure iterated until as many scores as possible were computed. Ratnaparkhi (1998) subsequently obtained similar results from an unsupervised method that required only part-of-speech tagging of the corpus, not parsing, by improving the heuristics by which the unambiguous training cases could be identified in the corpus.

Taking the four-element form of the problem, Brill and Resnik (1994) also obtained about 80 percent accuracy with a set of disambiguation rules that were derived from a corpus by means of the same supervised transformation-based learning method that Brill had earlier used for part-of-speech tagging. The rules state conditions under which a particular attachment is more likely: for example, 'the attachment point is the verb if the preposition is *in* and the noun of the prepositional phrase is a measure, quantity, or amount' or 'the attachment point is the object noun if the verb is a form of *to be*'. The semantic categorization of words for the rules (for example, characterizing a word as a measure, quantity, or amount for the rule mentioned above) is based on the WordNet electronic thesaurus (Fellbaum, 1998). Disambiguation initially assumes that the attachment is to the object noun. The rules are then applied, in a sequence of increasing specificity, to possibly change that; the provisional choice of attachment point might alternate several times as the rules are applied.

PRONOUN RESOLUTION

The ambiguity of pronouns (and anaphora in general) is different from the word sense ambiguity treated above in that there is no fixed inventory of candidate senses. Rather, the pronoun has an antecedent in the text with which it corefers; to disambiguate the pronoun is to find its antecedent, and the candidates are those elements of the preceding text that are 'available' for pronominal reference (in a sense that we will make more precise below). (*See* **Anaphora**; **Anaphora, Processing of**)

The antecedent of a pronoun is distinguished from its referent in that the antecedent is an element of text and the referent is an object in the world. Consider the following text from Charles Dickens's *Our Mutual Friend*:

> 'Let me', says the large man, trying to attract the attention of his wife in the distance, 'have the pleasure of presenting Mrs Podsnap to her host.'

The antecedent of *his* is the noun phrase *the large man*, and the referent of both is Mr Podsnap. The antecedent of a pronoun may be another pronoun; thus a text of the form *Mr Podsnap … He … He … He* (all about Mr Podsnap) creates a 'chain' of coreference, with the second and third pronouns each having the previous pronoun as its antecedent. Although one might have instead said that *Mr Podsnap* is independently the antecedent of all three pronouns, viewing antecedence as a chain appeals to our intuition that the antecedent of a pronoun must be recent within the text. (*See* **Story Understanding**)

The resolution of a pronoun is thus a two-stage process: determining the candidate antecedents, and then, if there is more than one, choosing among them. We will consider each stage in turn.

Candidates for Antecedence

While textual recency is a criterion for antecedence, it is neither necessary nor sufficient; indeed, there need be no single explicit textual antecedent. What matters most is that the referent be in the 'focus of attention' at the point at which the pronoun is uttered. That the antecedent need not be recent was shown by Grosz (1977) in her studies of people engaged in task-oriented dialogues, such as an instructor guiding an apprentice. Grosz found that when a partially completed task was resumed after a long intervening subtask, the speakers would often refer by pronouns to antecedents in the earlier discourse about the task; what mattered was that the particular task was again in the speakers' attention. That recency is not sufficient for an element to be available as an antecedent can be seen in this text:

> John put the wine on the table. It was brown and round.

Readers generally find this text to be somewhat odd, with the antecedent of *it* seeming to be *the wine*, even though *the table* is more recent and a table is more likely than wine to be brown and round. (In the terminology of centering theory, to be introduced below, this text is an example of a 'rough shift'.) That there need not be a single explicit antecedent can be seen in this text from Charles Dickens's *Our Mutual Friend*:

> Mrs Lammle bestowed a sweet and loving smile upon her friend, which Miss Podsnap returned as she best could. They sat at lunch in Mrs Lammle's own boudoir.

The antecedent of *they* is *Mrs Lammle* and *Miss Podsnap* together – a set that the reader must construct from separate elements of the text.

Choosing from Multiple Candidates

When there is more than one candidate antecedent, the choice among them is based on linguistic constraints and preferences and on common-sense knowledge of the world.

In most languages, pronouns are marked for gender, number, or both, and candidates that do not match these features are therefore immediately ruled out by these constraints; in English, a reference to a person can be eliminated as a possible antecedent for the pronoun *it*. Syntax also puts various restrictions on antecedence. For example, in English syntactic structures, if a nonreflexive pronoun functions as a complete noun phrase, its antecedent cannot be any node of the parse tree that is immediately dominated by another node that also dominates the pronoun. It is this rule that precludes *Nadia* being the antecedent of *her* in *Nadia baked her a cake*.

Syntactic structure can also determine a preference for one candidate over another. For example, a candidate antecedent that plays the same syntactic role as the pronoun is preferred over one that doesn't, especially if the sentences in which they occur exhibit 'syntactic parallelism'. For example, in *Nadia waved at Emily and then she shouted at her*, the preferred interpretation is that *she* is *Nadia* (both are subjects of their verb) and *her* is *Emily* (both are objects of their verb). Notice that if the pronouns are stressed heavily, the pattern is reversed; stress on a pronoun generally indicates that its antecedent is not the one that would normally be preferred (Kameyama, 1999).

One particularly influential theory of antecedence preference is centering theory (Walker *et al*., 1998). Centering theory relates the form chosen for a referring expression – such as the speaker's choice of whether or not to use a pronoun – to the focus of attention within the discourse, the syntactic structure of the text, and the difficulty of interpretation of the utterance. In the theory, each sentence within a discourse is said to have a 'center', which is, roughly, its topic or its most salient element; and each sentence of the discourse makes elements available, including its center, that could become the center of the subsequent sentence. These potential centers are ranked by their syntactic position – subject is ranked highest, then object, then other positions – and if any of these potential centers are indeed mentioned in the subsequent sentence, then the one that ranks highest in the first sentence is the actual center of the second sentence, regardless of its position in that sentence. Now, the center must always be pronominalized if any other element of the sentence is; thus, if a sentence contains just one pronoun and its antecedent cannot be found in the same sentence, that pronoun must be the center, and so its antecedent is therefore the highest-ranking potential center from the previous sentence. The following example is simplified from Thomas Bulfinch's *The Age of Fable*:

> Orpheus was presented by Apollo with a lyre and taught to play upon it. He did so to such perfection that nothing could withstand the charm of the music.

He is the only pronoun in the second sentence, so it must be the center, and its antecedent is *Orpheus*, which as subject of the first sentence outranks *Apollo*. This rule also explains the problem of the 'brown and round' example earlier; *it* is the only pronoun in the second sentence, and in the first sentence, *the wine* outranks *the table* as a potential center. Transitions like this, to a new center that is neither the center of the previous sentence nor its highest-ranking potential center, are very rare in naturally occurring text (Di Eugenio, 1998).

In observations such as these, centering theory thus provides a set of preferences that can be employed in pronoun resolution and an explanation of the difficulty that people experience when the expectations that these preferences entail are not fulfilled (Hudson-D'Zmura and Tanenhaus, 1998).

By applying constraints and preferences such those just described, a natural language system can often determine a unique antecedent for a pronoun. Systems differ in the exact rules that they apply, the order in which they apply them, and how they trade off conflicting constraints and preferences. The system developed by Lappin and Leass (1994), for example, achieved an overall success rate of 89 percent on within-sentence antecedence and 74 percent on cross-sentence antecedence; since within-sentence antecedence is more common, the overall success rate was 86 percent. But an error rate of 14 percent is still too high for most practical uses.

It is not surprising that a system using only syntactic constraints and preferences will make mistakes relatively often, as knowledge of what 'makes sense' is often required to choose the correct antecedent of a pronoun. Lappin and Leass give this example (from a computer manual):

> This green indicator is lit when the controller is on. It shows that the DC power supply voltages are at the correct level.

Their system incorrectly chooses *controller* over *green indicator* for *it*; the two alternatives are rated equally in all respects (each is the subject of its verb) except for recency, which favors *controller*. Clearly, *indicator* is, in general, a 'better' subject for the verb *show* than *controller* is; this suggests the use of selectional restrictions, as used for lexical ambiguity, as an additional constraint. An approximation to this, frequency of co-occurrence, is proposed by Dagan and Itai (1990): statistics gathered from a large corpus would be used to give preference to the candidate antecedent that occurs more frequently as the subject of the verb *show*. Incorporating this and other heuristics into a single process, Mitkov (1998) has achieved anaphor resolution accuracy approaching 90 percent.

References

Brill E and Resnik P (1994) A rule-based approach to prepositional phrase attachment disambiguation. In: *Proceedings, 15th International Conference on Computational Linguistics, Kyoto*, pp. 1198–1204.

Collins MJ (1996) A new statistical parser based on bigram lexical probabilities. In: *Proceedings, 34th Annual Meeting of the Association for Computational Linguistics, Santa Cruz, California*, pp. 184–191.

Dagan I and Itai A (1990) Automatic processing of corpora for the resolution of anaphora references. In: *Proceedings, 13th International Conference on Computational Linguistics, Helsinki*, vol. III, pp. 330–332.

DeRose SJ (1988) Grammatical category disambiguation by statistical optimization. *Computational Linguistics* **14**: 31–39.

Di Eugenio B (1998) Centering in Italian. In: Walker *et al*. (1998), pp. 115–137.

Fellbaum C (ed) (1998) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Grosz BJ (1977) *The Representation and Use of Focus in Dialogue Understanding*. PhD thesis, University of California, Berkeley, CA.

Hindle D and Rooth M (1993) Structural ambiguity and lexical relations. *Computational Linguistics* **19**: 103–120.

Hirst G (1987) *Semantic Interpretation and the Resolution of Ambiguity*. Cambridge, UK: Cambridge University Press.

Hudson-D'Zmura S and Tanenhaus MK (1998) Assigning antecedents to ambiguous pronouns: the role of the center of attention as the default assignment. In: Walker *et al*. (1998), pp. 199–226.

Johnson C and Fillmore CJ (2000) The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure. In: *Proceedings, 1st Meeting of the North American Chapter of the Association for Computational Linguistics, Seattle*, pp. 56–62.

Kameyama M (1999) Stressed and unstressed pronouns: complementary preferences. In: Bosch P and van der Sandt R (eds) *Focus: Linguistic, Cognitive, and Computational Perspectives*, pp. 306–321. Cambridge, UK: Cambridge University Press.

Kilgarriff A (1992) Dictionary word sense distinctions: an enquiry into their nature. *Computers and the Humanities* **26**: 365–387.

Kilgarriff A (1997) I don't believe in word senses. *Computers and the Humanities* **31**: 91–113.

Kilgarriff A and Palmer M (eds) (2000) *Computers and the Humanities*, **34**: 1–243. [Special issue on SENSEVAL.]

Lappin S and Leass HJ (1994) An algorithm for pronominal anaphora resolution. *Computational Linguistics* **20**: 535–561.

Lesk ME (1986) Automatic sense disambiguation using machine-readable dictionaries: how to tell a pine cone from an ice cream cone. In: *Proceedings, 5th International Conference on Systems Documentation, Toronto*, pp. 24–26. New York, NY: Association for Computing Machinery.

Mitkov R (1998) Robust pronoun resolution with limited knowledge. In: *Proceedings, 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Montreal*, pp. 869–875.

Ratnaparkhi A (1998) Statistical models for unsupervised prepositional phrase attachment. In: *Proceedings, 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics, Montreal*, pp. 1079–1085.

Resnik P (1998) WordNet and class-based probabilities. In: Fellbaum (1998), pp. 239–263.
Resnik P and Yarowsky D (1999) Distinguishing systems and distinguishing senses: new evaluation methods for word sense disambiguation. *Natural Language Engineering* **5**: 113–133.
Walker M, Joshi AK and Prince EF (eds) (1998) *Centering Theory in Discourse*. Oxford: Clarendon Press.
Yarowsky D (1992) Word-sense disambiguation using statistical models of Roget's categories trained on large corpora. In: *Proceedings, International Conference on Computational Linguistics, Nantes, France*, pp. 454–460.
Yarowsky D (1995) Unsupervised word sense disambiguation rivaling supervised methods. In: *Proceedings, 33rd Annual Meeting of the Association for Computational Linguistics, Cambridge, MA*, pp. 189–196.

Further Reading

Grosz BJ and Sidner CL (1986) Attention, intentions, and the structure of discourse. *Computational Linguistics* **12**: 175–204.
Hirst G (1981) *Anaphora in Natural Language Understanding: A Survey*. Berlin: Springer.
Ide N and Véronis J (eds) (1998) *Computational Linguistics* **24**: 1–165. [Special issue on word sense disambiguation.]
Jurafsky D and Martin JM (2000) *Speech and Language Processing*. Upper Saddle River, NJ: Prentice-Hall.
Mitkov R (2002) *Anaphora Resolution*. London: Longman.
Palmer M and Light M (eds) (1999) *Natural Language Engineering* **5**(2): i–iv and 113–218. [Special issue on semantic tagging.]
Resnik P (1999) Semantic similarity in a taxonomy: an information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research* **11**: 95–130.
Schütze H (1997) *Ambiguity Resolution in Language Learning*. Stanford, CA: CSLI.
Webber BL (1978) *A Formal Approach to Discourse Anaphora*. New York, NY: Garland.