4 Lexical disambiguation

There's no need to think, dear. Just do what's in the script, and it'll all come out right.
—Alfred Hitchcock[1]

4.1 What is necessary for lexical disambiguation?

The problem of determining the correct sense of a lexically ambiguous word in context has often been seen as one primarily of context recognition, a word being disambiguated to the unique meaning appropriate to the frame or script representing the known or newly-established context. For example, in the well-known SAM program (Cullingford 1978, Schank and the Yale AI Project 1975, Schank and Abelson 1977), each script has associated with it a set of word meanings appropriate to that script; in the restaurant script, there will be unique meanings given for such words as *waiter* and *serve*, and when (4-1) is processed:

(4-1) The waiter served the lasagna.

the fact that *serve* has quite a different meaning in the tennis script will not even be noticed (Schank and Abelson 1977: 183). (Certain words naming well-known people and things, such as *Manhattan*, are always present (Cullingford 1978: 13).)[2]

In its most simple-minded form, the script approach can easily fail:

(4-2) The lawyer stopped at the bar for a drink.

[1] Probably apocryphal.

[2] The script approach is in some ways reminiscent of early statistical approaches. In 1965, Madhu and Lytle proposed that in machine translation of scientific papers, it was generally sufficient for resolution of lexical ambiguity to know the likelihood of usage of each sense in the particular subfield of science that the paper was concerned with. The particular subfield of a paper was determined by looking at the subfield associations of the unambiguous words in it. Data for the probability of use in a particular sense in each subfield was obtained from textual analysis. A Bayesian formula then gave the probability of a given sense of a particular word in a particular context. 90% accuracy was claimed.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511553530.005
Cambridge Books Online © Cambridge University Press, 2009

If only the lawyering script is active, then *bar* as **an establishment purveying alcoholic beverages by the glass** will be unavailable, and its legal sense will be incorrectly chosen. If resolution is delayed until the word *drink* has had a chance to bring in a more suitable script, then there will be the problem of deciding which script to choose the sense from;[3] as Charniak (1984) points out, it is reasonable to expect in practice to have fifty or more scripts simultaneously active, each with its own lexicon, necessitating extensive search and choice among alternatives. Thus the main advantage of the script approach, having disambiguation "fall out" of context recognition, is lost.[4]

Further, even in a single script, a word, especially a polysemous one, may still be ambiguous. In the lawyering script, the word *bar* could mean **the physical bar of a courtroom** or **the legal profession**.[5] Moreover, choosing which script to invoke in the first place may require deciding the meaning of an ambiguous word, such as *play* in these sentences:

(4-3) Ross played with his toys. *(script = recreation)*

(4-4) Ross played his guitar. *(script = music-making)*

Note that the word *guitar* is by itself insufficient to invoke the music-making script. Compare:

(4-5) The baby played with the guitar. *(script = recreation)*

In general, word sense can depend not only upon global context, but also (or only) upon local cues, such as the meaning of nearby words. In (4-6):

(4-6) Nadia's car is a lemon-colored Subaru.

the **badly-made-car** meaning of *lemon* can be rejected in favor of the **citrus-fruit** meaning, without any consideration of global context; all one has to do is look at the word with which it has been hyphenated, knowing that color is salient and constant for only one of *lemon*'s two meanings. Often, all that seems necessary for disambiguation is that there be a "semantic association" between one sense of the ambiguous word and nearby words (Quillian 1962, 1969):

(4-7) The dog's bark woke me up. *(bark* ≠ **surface of tree***)*

[3] Schank and Abelson (1977: 183) claim that SAM would choose the non-script meaning of a word when the situation forces it, as with *serve* in sentence (i):

(i) The waiter served in the army.

but they are rather vague on how this happens.

[4] An additional argument against using context as the main source of information is the evidence that better readers depend LESS on context for word recognition than poor readers (West and Stanovich 1978; Stanovich, West, and Feeman 1981; West, Stanovich, Feeman, and Cunningham 1983; Stanovich, Cunningham, and Feeman 1984; Stanovich 1984; Nathan 1984).

[5] In fact, *Webster's ninth new collegiate dictionary* (Webster 1983) distinguishes seven different law-related senses of *bar*.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511553530.005
Cambridge Books Online © Cambridge University Press, 2009

As Hayes (1977a: 43) puts it, "an association between a sense of an ambiguous word and the context surrounding a use of that word is strong evidence that the interpretation of that word should come through that sense". The nearby disambiguating words may themselves be ambiguous; a well-known example (Small 1980) is *deep pit*. The word *deep* can mean **profound** or **extending far down** and *pit* can be **fruit stone** or **hole in the ground**; however, only one meaning of each fits with the other, so they are mutually disambiguating.

Knowledge about case slot flags and restrictions on case slot fillers is also a good source of disambiguating information. For example, (4-3) and (4-4) can be disambiguated easily if one knows that the **music-making** sense requires its PATIENT to be flagged by *OBJ*, while the **recreation** sense flags it with *with*. The **game-playing** sense also flags its PATIENT with *OBJ*, but can be distinguished from **music-making** by the fact that the former requires the PATIENT to be a game, while the latter requires a musical instrument.[6]

I noted in section 1.1.2 that the problem of deciding which case slot is flagged by a particular preposition or syntactic position is very similar to lexical disambiguation. The strategy of inserting pseudo-prepositions to represent case-flagging syntactic positions *(section 3.4)* makes the problems even more similar by making all case flags lexemes. The *play* examples suggest that verbs and case flags can be mutually disambiguating, just like the words in the *deep pit* example; though each may individually have many meanings, there will only be one combination of meanings that fits together (if the sentence is truly unambiguous).

Syntax can also supply disambiguating information. For example, the word *keep*, meaning **continue to do**, **continue to be**, or **maintain**, can be disambiguated by seeing whether its object is a gerund, an adjectival phrase, or a noun phrase:

(4-8) Ross <u>kept staring at Nadia's décolletage</u>.

(4-9) Nadia <u>kept calm</u> and made a cutting remark.

(4-10) Ross wrote of his embarrassment in <u>the diary</u> that he <u>kept</u>.

Sometimes, however, high-level inference, a relatively expensive operation, will be necessary. Compare:

(4-11) The lawyer stopped at the <u>bar</u> for a drink.

(4-12) The lawyer stopped at the <u>bar</u>, and turned to face the <u>court</u>.

Note that (4-12) has several sensible meanings: *bar* could be **the railing in a courtroom**, with *court* being the **judiciary assembled** in the courtroom, or it could be a **drinking establishment**, with *court* being a **courthouse** across the street, or a **tennis court**, or some other kind of court. Deciding which is better in a particular instance requires inference on the preceding context. Another example (from Hayes 1977a):

[6] This is, of course, a great simplification, made just to establish the point. In fact, *play* is a verb with a particularly complex and interesting case structure, analyzed extensively by Taylor (1975; Taylor and Rosenberg 1975).

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511553530.005
Cambridge Books Online © Cambridge University Press, 2009

(4-13) Nadia swung the hammer at the nail, and the head flew off.

The word *head* is most naturally interpreted as the **hammer-head**, not the **nail-head** (or Nadia's **person-head**), but figuring this out requires inference about the reasons for which one head or the other might have flown off, with a knowledge of centrifugal force suggesting the head of the hammer. In (4-14):

(4-14) Gag me with a spoon![7]

the disambiguation of *gag* (meaning either **choke** or **render unable to speak by blocking the mouth**) requires deciding which of its meanings might best be accomplished with a spoon, a non-trivial task.

Necessary for word sense disambiguation, then, are:

- a knowledge of context;
- a mechanism to find associations between nearby words;
- a mechanism to handle syntactic disambiguation cues;
- a mechanism to handle selectional restriction reconciliation negotiations between ambiguous words; and
- inference, as a last resort.

4.2 Lexical disambiguation research in AI

In this section I discuss work in AI on lexical disambiguation that takes into account local disambiguating cues. We will look at four very different approaches: those of Yorick Wilks, Philip Hayes, Branimir Boguraev, and Steven Small.

4.2.1 Wilks: Preference semantics

Wilks's Preference Semantics system (Wilks 1973, 1975b, 1975c, 1975d) was perhaps the first NLU system to be explicitly designed around the need for lexical disambiguation (Boguraev 1979). The system's strategy was based on selectional restrictions, expressed in the form of templates that meanings had to fit. However, the restrictions were not absolute, but rather expressed PREFERENCES only. A word sense that fitted the preferences was always preferred, but if none was available, the system would take what it could get. This permitted some metaphor to be handled in sentences like (4-15) (Wilks 1977, 1982b):

(4-15) My car drinks gasoline.

[7] Vogue expression, 1982–1983. COREY, Mary and WESTERMARK, Victoria. *Fer shurr! How to be a Valley Girl—Totally!* New York: Bantam, November 1982.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554360.005
Cambridge Books Online © Cambridge University Press, 2009

(*Drink* prefers an animate subject, but accepts a machine.) Not just verbs but all parts of speech were labeled with their preferences. For example, the adjective *big* was expected to qualify a physical object. Preferences were even used in anaphora resolution, as for the word *it* in (4-16) (Wilks 1975d), which may not refer to the rock because *drink* prefers a liquid object:

(4-16) I bought the wine, sat on a rock, and drank it.

This approach is consistent with the predictive method of parsing and semantic interpretation used by Riesbeck (1974, 1975; Schank, Goldman, Rieger and Riesbeck 1975; Riesbeck and Schank 1978); in Riesbeck's analyzer, a verb such as *drink* would set up an explicit prediction that the next thing in the sentence would be a liquid.

The shortcomings of this approach lie in the fact that preferences or selectional restrictions are, by themselves, inadequate to deal with all instances of lexical ambiguity, such as those that need global context, association of nearby concepts, or inference for their resolution. Boguraev (1979: 3.23–3.25) points out that Wilks's approach is also unable to deal adequately with polysemous verbs such as in the *keep* examples of section 4.1 ((4-8) to (4-10)), or these different senses of *ask*:

(4-17) Ross asked Nadia a question. (*ask* = **request an answer to**)

(4-18) Ross asked Nadia to come with him. (*ask* = **request [the performance of an action]**)

(4-19) Ross asked Nadia for the numbat. (*ask* = **request to be given**)

The problem, says Boguraev, is that templates simultaneously have to be both general enough to translate the input text to a semantic representation and specific enough to disambiguate words. For a discussion of the shortcomings of using preferences for anaphora resolution, see Hirst 1981a.

4.2.2 Boguraev: Semantic judgments

Boguraev's disambiguation system (1979) was based on Wilks's approach to semantics, but attempted to eliminate many of the shortcomings of the simple preference method by using not just preferences but SEMANTIC JUDGMENTS. His system integrated both lexical and structural disambiguation with semantic interpretation. (Boguraev's methods for structural disambiguation will be discussed in section 6.3.1.)

The system consisted of an ATN parser (*see section 1.3.2*) to which semantic procedures had been added as actions to be performed upon the completion of noun phrases and clauses. These procedures were responsible for both kinds of disambiguation and for building the semantic representation (in a formalism similar to Wilks's). They first applied selectional restrictions, preferences, verb case structures, and the like in an attempt to disambiguate words. Further lexical disambiguation was integrated with structural disambiguation; judgments were made

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009

on the SEMANTIC COHERENCE of potential readings, which were rejected if found implausible *(see section 6.3.1)*. If all readings were found implausible, this was signaled to the parser, and it would back up and try a different path. The semantic judgments were made solely on the basis of lexical knowledge; the system did not have any general world knowledge base.

4.2.3 Hayes: Disambiguation with frames and associations

Hayes's work (1976, 1977a, 1977b, 1978) on the CSAW program was perhaps the first to consider multiple sources of knowledge in lexical disambiguation, with an emphasis on finding semantic associations between words.[8] The program also used case structures and selectional restrictions, both absolute and preferential, for disambiguation. CSAW tried to use the most restrictive methods first: case slots, associations, and finally selectional restrictions. The program first parsed its input with an ATN, and then started disambiguating the words that had multiple senses, all of which were considered together, until it converged on a solution. If a choice occurred at any point, the possibilities were considered in parallel; multiple forking was permitted.

Association-finding was facilitated by the knowledge representation that Hayes used: a semantic network upon which a frame system was superimposed. Generally speaking, frame systems and semantic networks are simply notational variants for the same kind of structure, but in Hayes's representation both notations were used. One could think of the knowledge base as a system of frames, each frame being a small piece of semantic net; alternatively, one could think of it as a network divided up into areas, each area a frame. Frames participated in two hierarchies: one for the ISA relation, and one for the PART-OF relation. Noun senses were represented as nodes in the network, and verbs were represented as frames (Hayes 1976: 30).

Associations were then found by looking at this representation. For example, a node was associated with any frame it was in; the node representing a human hand was in the frame representing a person. Since Ross is an instance of a person, a disambiguating association was found in (4-20):

(4-20) Ross's <u>hand</u> was covered with bandages.

A node was also associated with any frame that was an ancestor or descendant in the ISA hierarchy of the frame it was in, and other associations were specified by rules operating on the ISA and PART-OF hierarchies (Hayes 1976: 126). Such

[8] However, Quillian's Teachable Language Comprehender (1969) included a rudimentary form of disambiguation by association. Given (i):

(i) John shoots the teacher.

if John was known to be a photographer or a gangster, then word association would be used to determine whether *shoot* meant **photograph** or **assault with a firearm**. Quillian first proposed the technique in 1961 (see below, and Quillian 1962).

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009

associations relied on finding a CHAIN OF CONNECTIONS in the knowledge base between a node and a frame.

Hayes found that association is useful mostly for dealing with homonyms, where there are large differences between the senses; it is less successful with polysemous words (Hayes 1976: 30), where an association for more than one sense will tend to be found. Although verbs as well as nouns can be disambiguated by association, Hayes's treatment did not include verbs (1977: 46) because of their extra complexity. (Recall also that verbs tend to be polysemous while nouns tend to be homonymous—see section 1.1.2.)

4.2.4 Small: Word expert parsing

The Word Expert Parser of Steven Small (1980, 1983; Small and Rieger 1982) featured an unconventional architecture in which each word in the lexicon was represented by a procedure, and the parsing, disambiguation, and semantic interpretation of a sentence were performed by a consortium of the procedures that represented the words of the sentence. As the system's name suggests, each procedure could be thought of as an expert on both the syntax and semantics of the word that it represented. Together, these word experts built a semantic representation of the input, performing disambiguation and minimal syntactic analysis on the way. A distributed control environment governed the experts, which ran as deterministic coroutines, letting each have its turn when appropriate.

Each expert for an ambiguous word included a DISCRIMINATION NET for the possible meanings of that word. By asking appropriate questions of context and of its environment (that is, of the experts for nearby words), the expert traversed the net to reach a unique meaning for its word. The expert would report partial results as it became sure of them, and would go to sleep if it needed to await results that other experts hadn't yet determined. When it determined its correct sense, the expert added the result to the conceptual structure that was being built as the representation of the sentence.

Because the system conflated parsing, disambiguation and semantic interpretation, each expert had a lot of work to do, and experts were the sole repository of linguistic knowledge in the system.[9] Word expert procedures were therefore large and complicated and necessarily individually hand-crafted; "the construction of word experts requires patience, dedication, and finesse" (Small 1980: 200). For example, the expert for *throw* is "currently six pages long, but should be ten times that size" (Small and Rieger 1982: 146). This is the biggest drawback of the Word Expert Parser: its rejection of general mechanisms in favor of a very large number of loosely constrained particular ones, loosely related to one another.[10]

---

[9] This is in accord with Small's view that language is too irregular for generalizations to be usefully captured at any level higher than the lexical (Small and Rieger 1982: 90); *cf.* footnote 2 of chapter 1.

[10] To some large degree this is a religious argument. If one accepts Small's view of language, then

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511553868.004
Cambridge Books Online © Cambridge University Press, 2009

4.3 Psycholinguistic research on lexical disambiguation

There has been much psycholinguistic research in the last decade on how people deal with lexical ambiguity and choose the correct word sense. In this section I look at this research, in preparation for developing an AI lexical disambiguation mechanism *(cf. section 1.4)*.

In general, people do not notice occurrences of lexical ambiguity, and seem to disambiguate without any conscious effort.[11] There are exceptions to this, of course. Many jokes rely on the perceiver seeing only one meaning at first, and then noticing the other (Hirst 1988):

(4-21) [From the television program *The two Ronnies*]
Ronnie Corbett: In the sketch that follows, I'll be playing a man whose wife leaves him for the garbageman.
Ronnie Barker: And I'll be playing the garbageman who refuses to take him.[12]

(4-22) **Julia:** Why didst thou stoop then?
Lucetta: To take a paper up that I let fall.
Julia: And is that paper nothing?
Lucetta: Nothing concerning me.
Julia: Then let it lie for those that it concerns.
Lucetta: Madam, it will not lie where it concerns,
Unless it have a false interpreter.[13]

while others, puns and double entendres, rely on just the opposite, that both meanings be seen:

(4-23) One man's Mede is another man's Persian.[14]

(4-24) There is no <u>seeming</u> mercy in the King.[15] *(seeming = both* **apparent** *and* **fraudulent***)*[16]

---

his approach is both reasonable and necessary, and the Word Expert Parser is a good first attempt at a system based upon this view. My opinion is that this is a rather inconvenient religion, and not one to be adopted lightly. Before one dons a hair shirt, one has to be convinced that God really thinks it necessary.

[11]Nevertheless, Mohanty (1983) has shown that the presentation of a lexically or structurally ambiguous sentence causes a small but significantly greater increase in the rate of a person's heartbeat than an unambiguous sentence does.

[12]DEYKIN, Graham. In VINCENT, Peter (compiler). *The two Ronnies—but first the news*. London: Star, 1977. 36.

[13]SHAKESPEARE, William. *Two gentlemen of Verona*. 1594. I, ii, 69–75.

[14]COHAN, George M. Attributed.

[15]SHAKESPEARE, William. *Henry IV, part I*. 1590. V, ii, 34.

[16]"Worcester is withholding the King's offer of amnesty from Hotspur because he believes it fraudulent, merely a seeming mercy, and this gives his words a negative dramatic irony. They gain also a positive irony from the audience's belief that the King's offer is sincere; Worcester is more right than he knows, since the King's mercy is not seeming but genuine" (Mahood 1957: 43).

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009

(See Hirst 1988 for more discussion; see Hillier 1974 for a discussion, with examples, of the ten different types of pun; see Raskin 1985 for a discussion of linguistic ambiguity in humor; see Levine 1965, 1985 for discussions of the use and benefits of other kinds of deliberate linguistic ambiguity in non-Western cultures.)

Shamelessly leaving aside such awkward counterexamples, we can consider how people might disambiguate words. The main problem we look at is the LEXICAL ACCESS question:

- Do people consider (unconsciously) some or all of the possible meanings of an ambiguous word, or do context and/or expectations take them straight to the "correct" meaning?

The competing hypotheses are these:

- The PRIOR CHOICE hypothesis: "Prior context biases the interpretive process before the ambiguity itself is encountered . . . perhaps allowing only a single reading ever to be accessed" (Swinney and Hakes 1976: 683).
- The ALL-READINGS hypothesis: "Prior context influences the interpretation of an ambiguity only after all readings are accessed; it aids in the selection of an appropriate reading to retain from those accessed, but does not affect the access process itself" (Swinney and Hakes 1976: 683).
- The ORDERED SEARCH hypothesis: "The order of search [of senses of an ambiguous word] is determined by frequency of usage, the most frequent being first. The search is self-terminating, so that as soon as an acceptable match occurs, no [other] entries will be checked" (Hogaboam and Perfetti 1975: 66).

Another question that immediately arises is the DECISION POINT problem:

- If more than one meaning is accessed, how and when is a choice made?

There are three possibilities: that the choice is virtually immediate; that it does not happen until the end of the clause (or some smaller syntactic unit), with the several meanings remaining around until then; and that it happens as soon as enough information is available, whether this be immediately or later.

The discussion below will deal with the lexical access question rather than the decision point question. This is because research has focused so far almost exclusively on cases where both enough information is present at the time of the word's occurrence for it to be immediately disambiguated and the word is the last of the clause anyway.[17] It has been found, unsurprisingly, that disambiguation is immediate in such cases; in addition, the work of Swinney (1979) and Onifer and Swinney (1981) that I will discuss in section 4.3.4 suggests that this result also holds when there is sufficient information, even if the word is not clause-final. Further, intuition suggests that ambiguities are always resolved by the end of the sentence, with a good guess being made if the information provided is insufficient.

[17] A notable exception is the work of Carpenter and Daneman (1981).

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009

The review below will not attempt to cover the large body of literature, but rather a representative sample. Moreover, I will be almost exclusively concerned with homonymous nouns; there seems to have been very little psycholinguistic research on other forms of lexical ambiguity, except for a bit on homonyms that are also noun–verb categorially ambiguous. I first discuss the concept of spreading activation, which will be necessary for the subsequent discussion of the three hypotheses for the lexical access question.

4.3.1 Semantic priming and spreading activation

A phenomenon whose psychological reality is well established is SEMANTIC PRIMING: the fact that the mental processing of a particular concept will facilitate processing of other concepts that have a semantic relationship to the first. For example, Meyer and Schvaneveldt (1971) found that subjects could answer the question "Are these strings both English words?" faster if the stimulus was a related pair like *doctor–nurse* than an unrelated pair like *doctor–butter*. Accessing the first word seemed to facilitate access to the second when it was semantically related.

Models of semantic priming are generally based on SPREADING ACTIVATION. It is assumed that the mental representation of concepts is a network of some kind, similar to a semantic network or network of frames, in which semantically related concepts are close to one another. Using a concept in the network causes it to become ACTIVATED; for example, processing a word will cause the word's meaning to become activated. This activation, however, is not limited just to the concept accessed. Rather, activation will spread from the ORIGIN to nearby nodes, causing them to also be activated. When a node is activated, access to it is facilitated. Thus, seeing the word *doctor* activates the **doctor** concept and also nearby, semantically related concepts such as **nurse**. When the word *nurse* is then seen, its corresponding node has been pre-activated, making access to it faster than access to concepts like **butter** that have not been activated.

Spreading activation models generally assume that the more a concept is used, or the more activation that is spread to it, the more highly activated it becomes. For example, Reder (1983) found that priming a concept twice doubled its degree of facilitation. However, activation gets weaker as it spreads until it can spread no further, so that the degree of activation of a concept will be a function of its semantic closeness to the origin. In addition, concepts do not remain activated long; their activation decays with time until they return to their normal state.

One of the earliest models of spreading activation was that of Quillian (1962, 1967, 1968, 1969), who, following early work on semantic networks by Richens (1958) and Masterman (1961), developed a semantic network in which connections between concepts could be found. This work inspired considerable research in cognitive psychology on spreading activation models; for an overview, see Collins and Loftus 1975 and Lorch 1982, or Anderson 1976, 1983.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009

In the sections below, we will see semantic priming and spreading activation used both as tools for the investigation of lexical disambiguation and as explanations of the disambiguation mechanism itself.

> *Wigan* (n.) — If, when talking to someone you know has only one leg, you're trying to treat them perfectly casually and normally, but find to your horror that your conversation is liberally studded with references to *(a)* Long John Silver, *(b)* Hopalong Cassidy, *(c)* the Hokey Cokey, *(d)* 'putting your foot in it', *(e)* 'the last leg of the UEFA competition', you are said to have committed a *wigan*.
>
> The word is derived from the fact that sub-editors at ITN used to manage to mention the name of either the town Wigan, or Lord Wigg, in every fourth script that Reginald Bosanquet was given to read.
>
> —Douglas Adams and John Lloyd[18]

4.3.2 Phoneme monitoring and prior choice activation models

An answer to the lexical access question might be based on the observation that if the prior choice hypothesis is correct, then ambiguous words should require no longer to process than other words, as their ambiguity will not affect their processing time. This would be predicted by the script approach of Schank and Abelson (1977), a strong form of the prior choice hypothesis in which the invocation of a script gives the system a list of unique word meanings for the present context *(see section 4.1)*. Thus any extra processing time incurred by lexical ambiguity happens when scripts are initially invoked, rather than when individual word occurrences are processed.

The data on the processing times for ambiguous words are somewhat equivocal.[19] Foss (1970) asked subjects to press a button upon hearing a specified phoneme in a sentence; reaction time was found to be greater if the target phoneme was immediately preceded by an ambiguous word, suggesting that lexical processing is slowed down by the ambiguity and that ambiguous words therefore require greater processing. Foss and Jenkins (1973) found that this effect was present even when the sentence provided a context that biased the meaning choice:

(4-25) The farmer put his <u>straw</u> beside the machine. *(straw* ≠ **drinking-straw***)*

The implication is that despite the context, both meanings of the word were retrieved and considered. However, a replication with the phoneme-monitoring task by Swinney and Hakes (1976), in which the bias of the context was extremely

---

[18] ADAMS, Douglas and LLOYD, John. *The meaning of Liff.* London: Pan Books and Faber and Faber, 1983. Reprinted by permission.

[19] The content of this paragraph and the next is based on Foss and Hakes 1978: 120–124. A good overview and critique of the work discussed in the remainder of this section is also given by Simpson (1981).

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009

strong, found no difference between the reaction time for ambiguous and unambiguous words:

(4-26) Rumor had it that, for years, the government building had been plagued with problems. The man was not surprised when he found several spiders, roaches, and other bugs in the corner of his room. *(bug ≠* **hidden microphone***)*

These data suggest a partial prior choice model based on activation of word senses. When the correct sense for a particular occurrence of a word is sought, only the most active senses are considered. If the preceding context has highly activated one sense, only that sense will ever be seen; if there is no contextual bias, all senses will be seen, and a choice made from them.[20] A mild contextual bias will not activate a single meaning enough to prevent some of the other meanings from being seen.[21] Blank and Foss (1978) similarly found that the phoneme-monitoring task was facilitated by semantically related prior context, even if the word at which the target occurred was not ambiguous.

This model predicts that if, for some reason, the incorrect meaning of an ambiguous word has been highly pre-activated, then that word will be misinterpreted, resulting in a SEMANTIC GARDEN-PATH SENTENCE. This does in fact occur; most people have a great deal of trouble with NEGATIVELY PRIMED sentences such as these:

(4-27) The astronomer married the star.

(4-28) The sailor ate the submarine.

(4-29) The watchmaker removed the tick.

(4-30) The rabbi who addressed the congregation was hit on the temple.

(4-31) The catcher filled the pitcher.[22]

---

[20] Note that it is necessary to assume that spreading activation does NOT spread through the mental representation of words themselves, but only through the representation of meanings. Otherwise, activation would spread from one sense of a word through the word itself to another of its senses, preactivating more than one of its meanings! Lucas (1983, 1984) tested activation of the various senses of ambiguous words just at their onset, and found that only the sense primed by the context was active, so the assumption seems to be a safe one.

[21] This model is similar to Morton's "logogen" model (1969).

[22] Vocabulary for the non–North American reader: A *submarine* is a type of sandwich, and a *catcher* and a *pitcher* are players on a baseball team; readers who prefer cricket to baseball may substitute (i) for (4-31):

(i) The wicket-keeper wore the bowler.

Example (4-27) is due to John Anderson (Lynne Reder, personal communication), and examples (4-29) to (4-31) are based on ones from Reder 1983; (4-28) and (i) are my own.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554360.006
Cambridge Books Online © Cambridge University Press, 2009

Reder (1983) found that comprehension errors increased markedly in negatively primed sentences. For example, in (4-27), people often take *star* to be **astronomical object** instead of **celebrity**, although this violates selectional restrictions on *marry*, and then become confused, or attempt to reinterpret the sentence metaphorically. (There seem to be wide individual differences in the processing of such sentences. This is to be expected to some extent, as people will differ in their mental organization of concepts and hence in the exact boundary effects of spreading activation.)

Note that the partial prior choice model differs significantly from the script model. While it claims that there is often a prior choice, it does not rely solely on context to provide a unique word sense, but also contains a complete (unspecified) disambiguation mechanism for use in mild or neutral contexts where no single sense is more highly activated than the others. Also, it accounts not only for the effect of global context but the effect of local word association as well; scripts, even if they had psychological reality, could not be the only sense pre-selection mechanism.

There are, however, three major criticisms of these results. The first is that none of these experiments controlled for relative sense frequency.[23] The ordered search hypothesis suggests that the word *bug* in (4-26) is disambiguated quickly just because **insect** is the more common meaning; most people would require only one lexical access to get this sense, whereas a context indicating the **hidden microphone** sense would require a second access after the **insect** sense failed. On the other hand, (4-25) would require an average of 1.5 accesses to disambiguate *straw*, as neither of its senses dominates the other, and we would therefore expect that about 50% of the subjects would get it on the first access and 50% would require a second access.[24] Thus the discrepancy between the results of Foss and Jenkins and those of Swinney and Hakes would be explained as artifacts of their experimental materials.[25]

---

[23]Experiments on eye fixations in reading suggest that lexical retrieval time is in general a function of word frequency (Just and Carpenter 1980).

[24]Foss and Jenkins's own data on meaning preferences for *straw* show that in a neutral context (*The merchant put the straw beside the machine*), 48% of subjects chose the **hay** meaning and 52% chose **drinking-straw**.

[25]Also indicating an effect of word frequency are the results of Carpenter and Daneman (1981), who had subjects read aloud texts in which one of the meanings of a heteronym (an ambiguous word whose meanings have different pronunciations) was biased by the previous context. They found that readers not infrequently said the more frequent meaning of the word, even when the bias was otherwise, and did not always correct themselves even when the rest of the sentence made it clear that their choice was wrong.

One cannot generalize from their results, however, as reading aloud may require different strategies from normal sentence comprehension, since there is not the same motivation to process the sentence completely, but instead a desire to say the word as soon as possible after seeing it (though voice trails a word or two behind eye fixation), and simple word recognition without complete comprehension is

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.006
Cambridge Books Online © Cambridge University Press, 2009

The second criticism of the phoneme-monitoring experiments is that they infer the difficulty of processing an ambiguous word from the degree to which it seems to interfere with the phoneme recognition process. However, the target phoneme was always in a word after the ambiguous word, leaving a relatively long temporal gap after the occurrence of the ambiguity. Thus, phoneme-monitoring tasks at best give information about the processing load AFTER lexical access, but provide no clue about lexical access itself (Swinney 1979). In the Foss (1970) and Foss and Jenkins (1973) experiments, the target phoneme usually started the syllable after the ambiguous word, but was sometimes the start of the second syllable following; in Swinney and Hakes's (1976) experiment, the target was as many as three syllables away from the ambiguity ("never more than two words" (Swinney and Hakes 1976: 685)). This alone may account for the difference in the results, as Cairns and Kamerman (1975) have shown that a two-word distance between the ambiguity and the target is sufficient to eliminate any latency effect of the ambiguity.

The third criticism is that the length and phonological properties of the words used in the tests could account for most of the effect found. See Newman and Dell 1978 or Swinney 1979: 647 for details of this point, which need not concern us further here.

I now turn to work that attempts to overcome these criticisms. In the next section, I look at work that controls for dominant word sense, and in the section after that, at work that avoids the other methodological problems of phoneme monitoring.

4.3.3 The ordered search hypothesis

Hogaboam and Perfetti (1975) tested the ordered search hypothesis with a set of ambiguous words that each had a dominant sense and a secondary sense;[26] for example, the dominant sense of *arms* is **human limbs**, and the secondary sense is **weapons.**[27] Instead of using a phoneme monitoring task, they asked subjects to respond as rapidly as possible when the last word of a sentence could have a meaning other than the one it had in the sentence. Given (4-32):

(4-32) A good sprinter uses his arms.

a subject would have to respond positively, and then briefly explain the **weapons** sense of *arms*. The ordered search hypothesis predicts that subjects will be faster

---

adequate for giving the pronunciation of most words; *cf.* Shallice, Warrington, and McCarthy 1983.

[26]There is evidence that the most frequent sense is psychologically distinguished. For example, schizophrenia patients often interpret ambiguous words only with their most frequent meaning, even if the result is nonsense (Chapman, Chapman, and Miller 1964; Benjamin and Watt 1969).

[27]These norms were determined by word association tests. Given a stimulus such as *arms*, most subjects would respond with, say, "legs", while few would say "rifles" or "right to keep and bear shall not be infringed".

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009

on this task when the sentence uses the secondary meaning, as the primary meaning will have already been accessed and rejected, and the subject will know the word to be ambiguous. On the other hand, if the dominant meaning is used, the subject will still have to see if a second access is possible, in order to determine whether the word is ambiguous, and reaction time will be slower. Both the prior choice and all-readings hypotheses predict no difference in the two cases. Hogaboam and Perfetti found the result predicted by the ordered search hypothesis. However, as Swinney and Hakes (1976) and Onifer and Swinney (1981) point out, the nature of the task, in which subjects explicitly and consciously listened for ambiguity, may have affected the manner in which they processed the sentences, and thus the results may not reflect normal unconscious lexical access.[28] Simpson (1981) therefore tested the same hypothesis, but used a simple lexical decision task in which an ambiguous word was followed by a second word, either related to its dominant meaning, or to its secondary meaning, or completely unrelated. Supporting the ordered search hypothesis, he found that response time on the second word was faster only when it was related to the dominant meaning of the first word; words whose relationship was to the secondary sense had a reaction time as slow as unrelated words.

However, Simpson repeated his experiment with the addition of context; the ambiguous word was now the final word of a sentence that gave either a strong bias, a weak bias, or no bias to the ambiguous word. In the unbiased case, the results were as before. However, in a strongly biased context, retrieval was facilitated only for words related to the biased sense of the ambiguous word, regardless of whether it was the dominant or secondary sense. In a weakly biased context, a dominant sense facilitated retrieval only of words related to that sense, but a secondary sense facilitated both. Simpson's interpretation of these results is that the dominance of one sense and the effects of context are independent. Without context, dominance prevails, and only the primary sense is initially retrieved. In a strong context, context effects prevail, and only the appropriate sense is retrieved initially. But when the context only weakly chooses the secondary sense, neither context nor dominance prevails, and both senses are retrieved. The activation model can account for these results if we allow two changes: first, that dominant word senses are given a "head start" in the activation race, a lead that can be overcome by strong bias to the secondary meaning but only equaled by weak bias; and second, that if the sense or senses retrieved the first time are found unacceptable, the other senses are then retrieved.

[28] An example of how the nature of the task can affect results, even when the subjects are not conscious of ambiguity in the stimuli: Yates (1978) had subjects respond as fast as possible with the truth of statements of the form *An A is a B*, where *A* was a homonym and *B* was either its dominant sense or a very unusual sense: *A bug is an insect*, *A bug is a Volkswagen*. If subjects received trials that used dominant and unusual senses equally often, then semantic priming facilitated access to both senses; if they received trials that mostly used dominant senses, then priming facilitated dominant senses more than it did secondary senses. Note that subjects were not told about the ambiguity in the stimuli, nor about the distribution of dominant and secondary senses that they could expect.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511553530.005
Cambridge Books Online © Cambridge University Press, 2009

4.3.4 Semantic priming and the all-readings hypothesis

Contrary to Simpson's results, other recent work provides evidence for the all-readings hypothesis. Swinney (1979) used the test materials of the Swinney and Hakes (1976) experiment but substituted a lexical decision task ("Is this string a word?") for phoneme monitoring. To reduce the temporal gap, the string for the lexical decision was presented on a computer terminal exactly as the end of the ambiguous word was heard. He found that the speed of the lexical decision task was increased when the string was a word related to some sense of the ambiguous word, even if the context was strongly biased toward another sense. This was true even for strings related to the secondary sense when the context was biased to the dominant sense (Onifer and Swinney 1981). The implication is that both senses were accessed and primed in all contexts. A similar experiment by Tanenhaus, Leiman, and Seidenberg (hereafter *TLS*) (1979) showed that the effect held even for categorially ambiguous homonyms;[29] for example, a sentence such as (4-33):

(4-33) They needed a new <u>sink</u>.

facilitated the word *swim*, which is related to the verb *sink* but not to the noun form used in the stimulus.

When the experiment was repeated with the lexical decision task occurring three syllables after the ambiguous word (in the replication by TLS), or 1.5 seconds after it (in the Onifer and Swinney 1981 replication), only strings related to the contextually appropriate meaning of the ambiguous word were facilitated, suggesting that the activation of the inappropriate sense decays, or is actively suppressed (TLS 1979: 436), in less than the span of three syllables (or 1.5 seconds). In fact, TLS found that with categorial ambiguities, the facilitation effect had disappeared within only 200 milliseconds; Lucas (1983, 1984) found it still present after 100 and 125 milliseconds with noun–noun ambiguities.[30]

However, a subsequent study cast doubt on the generality of these results. In a series of experiments, Seidenberg, Tanenhaus, Leiman, and Bienkowski *(STLB)* (1982; Seidenberg and Tanenhaus 1980) varied the type of ambiguity—noun homonyms or categorially ambiguous noun–verb homonyms—and the type of context that biased the sense of the ambiguous verbs. The types of context were as follows:

- Neutral—No contextual cues to resolve ambiguity: *Joe buys the <u>straw</u>.*

[29]Instead of a lexical decision task, TLS's experiments used word-naming; the strings were always words, which the subjects had to say out loud, thereby stopping a sound-activated timer.

[30]An initial explanation for the discrepancy between Simpson's results *(see previous section)* and TLS's was that Simpson had used a 120-millisecond gap between the stimulus and target (1981: 134); however, Lucas's results cast doubt on this. Nevertheless, these every-millisecond-counts results should be treated with caution. Simpson points out (1981: 130) that determining the exact millisecond at which the stimulus ends and timing should commence is made extremely difficult by variations in the characteristics of the final phoneme of the stimulus, the volume of the speaker's voice, and so on.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009

Table 4.1. *Summary of results of Seidenberg et al (1982)*

TYPE OF CONTEXT	TYPE OF LEXICAL AMBIGUITY	OUTCOME
Neutral	Noun–Noun	Multiple access
Syntactic	Noun–Verb	Multiple access
Semantic priming	Noun–Noun	Selective access
Semantic priming	Noun–Verb	Multiple access
Non-priming biased	Noun–Noun	Multiple access

(From Seidenberg, Tanenhaus, Leiman, and Bienkowski 1982.)

- Semantic priming—The context contains a word associated with one sense of the ambiguous word: *The bridge player trumped the spade.*
- Syntactic bias—The syntax of the context selects the noun or verb sense of a categorially ambiguous word: *The congregation rose.*
- Non-priming biased—The context selects one sense of a noun homonym, but does so without containing any word related to that sense: *The man walked on the deck.*

Using the same techniques as the previous experiments, they found that all senses seemed to be activated (for less than 200 milliseconds), EXCEPT in the case of noun homonyms in a semantically primed context, when only the context-related sense was activated;[31] table 4.1 summarizes these results. Thus prior choice occurs only when the context contains a semantically related concept; a context that constrains the sense only by selectional restrictions or the like does not inhibit multiple access. Further, semantically primed contexts do NOT inhibit multiple access in categorially ambiguous words.[32]

[31] Lucas (1984) has suggested that there is multiple access in ALL cases, and in apparent exceptions it is simply extremely fast.

[32] In contrast to this finding, Ryder and Walker (1982) found that without context only the most frequent sense of a categorial ambiguity seemed to be activated. It is possible that this was because the ambiguous word was presented for a full second before the probe; STLB's results suggest that this was long enough for a meaning to have been activated and decay again. On the other hand, even though they had a 500-millisecond delay after the ambiguous stimulus, Oden and Spira (1983) obtained results partly supporting STLB: both senses of a categorially ambiguous word were activated, though the one to which the context was biased was activated more. Unfortunately these results must be used with caution, as there were errors in their experimental materials. For example, the words *bounce* and *jump* were used to bias the word *spring* toward the sense **coil**, and the senses of the homophone pair *die* and *dye* were taken to be about equally common, although they clearly are not.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511553486.005
Cambridge Books Online © Cambridge University Press, 2009

Swinney (1979), Onifer and Swinney (1981), and Simpson (1981), although they used strongly biased contexts, did not control for semantic priming as the source of the bias; some of their sentences contained semantically related words, while others had non-priming bias (Onifer and Swinney 1981:234):

(4-34) The team [*prime*] came out of the locker room with fire in their eyes after the coach delivered what was perhaps the best speech of his life.

(4-35) John sat down to make out checks for the monthly bills, but could not find a single working pen anywhere in the house.

STLB suggest that this may be why Swinney and Onifer and Swinney found apparent multiple access in all cases.[33] Not controlling for this type of bias may also have confounded Simpson's results.

STLB suggest a model in which semantic priming and frequency of use both affect activation, but non-priming bias does not. If one sense is much more preactivated than the others, it alone is selected; otherwise several or all senses are. Also in this model, categorially ambiguous words are represented separately for each syntactic category; thus semantic priming of one noun sense of a word may cause that sense to be favored over other noun senses, but won't restrict access to any verb senses of the same word.[34]

Subsequent work has sought to clarify the role of word frequency in this model. Simpson and Burgess (1985; Simpson 1984), presenting homographs out of context, measured facilitation at several points after the stimulus, from 16 milliseconds to 750 milliseconds. They found the dominant sense active at all points; the secondary sense became active more slowly, however, being facilitated between 100 and 500 milliseconds, but not at 16 or 750 milliseconds. This suggests that dominant senses have a "head start", but not so as to preclude activation of a secondary sense; in the absence of context, the dominant sense wins, and the activation of the other sense decays.

4.3.5 Conclusion

There are clearly many questions yet to be resolved in the study of human lexical access and disambiguation. However, this much seems clear: in many cases, more than one meaning of an ambiguous word is accessed. Semantic priming and frequency of a particular sense can facilitate lexical access and disambiguation, and

[33] They also suggest that the difference may be due to the fact that the priming word, when there was one, was further from the ambiguity in Swinney's and Onifer and Swinney's experiments—four or more words earlier, instead of two or three as in STLB's sentences, with the priming effect already starting to decay. However, if priming has an effect at all, one would expect it to last longer than that, or it would lose much of its functional advantage. A more interesting question is what the effect of a clause boundary between the prime and the ambiguity might be.

[34] See also Stanovich and West 1983b for a model of contextual priming that includes spreading activation.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009

in some cases may cause one meaning to be accessed to the exclusion of others. In the next chapter, I develop a lexical disambiguation system for use with Absity that has similar properties.

Downloaded from https://www.cambridge.org/core. University of Toronto, on 12 Oct 2018 at 13:30:52, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/CBO9780511554346.005
Cambridge Books Online © Cambridge University Press, 2009