Representation and processing of lexically ambiguous words

Stephen J. Lupker
University of Western Ontario

John and Pat looked over at their friend Bob. Both of them sighed. John turned to Pat and said, "You know, Bob could use a belt".

Readers of this passage are confronted with a basic problem common not only in English but also in many other languages. Words are inherently ambiguous in that most have multiple legitimate meanings. What does Bob actually need here, something to hold his pants up? A good smack? A stiff drink? Or, perhaps Bob is a down-on-his-luck developer who could get back on his feet if he could just get his hands on a patch of undeveloped land. In the end, context will help readers determine which of these things Bob's friends think he needs. Without context, however, one, some or all of these possible meanings may have been activated and considered by readers. It is that process, the process by which readers (and listeners) activate meanings and, ultimately, resolve ambiguities that is the topic of the present chapter.

Much of the work on this topic, particularly, prior to 1990, was directed at the questions of: 1) whether all meanings of an ambiguous word were initially activated even if context favored only one (Glucksberg, Kreuz & Rho, 1986; Onifer & Swinney, 1981; Simpson, 1981; Swinney, 1979; Seidenberg, Tanenhaus, Leiman & Bienkowski, 1982; Tabossi, 1988; Tabossi, Colombo & Job, 1987; Tanenhaus, Leiman & Seidenberg, 1979) and 2) whenever more than one meaning was activated, whether the more dominant meaning was activated more rapidly (Hogaboam & Perfetti, 1975; Simpson, 1981; Simpson & Burgess, 1985; Swinney, 1979). Much of this work involved priming paradigms. Essentially, ambiguous words were presented in sentences with or without preceding, disambiguating context. Latencies to respond to words related to one or the other meaning of the ambiguous word were then evaluated.

Some of these studies are reviewed at the end of the chapter. The main focus of the chapter, however, is the impact of ambiguity on word processing out of context, mostly, in single word presentation tasks. These tasks should give us the best chance of uncovering the basic principles underlying how these words are represented in memory and how that memory representation is accessed. (Also see Rastle, this volume, for a discussion of the word recognition literature based on these types of experiments.) A discussion of the key phenomenon is followed by a discussion of attempts to model that phenomenon based on some possible ways that ambiguous words might be represented in memory. Next, more recent attempts to distinguish between different types of ambiguous words are discussed and what that research might have to say about the representation question. Finally, we examine a prediction of all current models by considering data from other, neutral-context tasks before turning to the question of the impact of context.

12.1 The ambiguity advantage

12.1.1 Rubenstein and colleagues

Intuitively, the expectation most people have is that ambiguous words, like *belt*, would be harder to process than unambiguous words. Indeed, most models of the process implicitly or explicitly make exactly that prediction. The story, however, is more complicated. Ambiguity can also have benefits depending on what the reader is attempting to do with the word.

Rubenstein, Garfield and Millikan (1970) appear to have been the first researchers to compare the processing of ambiguous versus unambiguous words (or, in their terms, homographs versus nonhomographs). The homographic status of each word was derived from a subjective rating procedure carried out by 20 independent raters. The experimental task was a lexical decision task. The results were straightforward. Lexical decisions were faster to ambiguous words than to unambiguous words matched on frequency and concreteness. Subsequently, Rubenstein, Lewis and Rubenstein (1971) reported that this ambiguity advantage was essentially restricted to those ambiguous words that had two, unrelated, equiprobable meanings. That is, ambiguous words like *glue*, which means both “a strong adhesive substance” and “to cause to stick tightly with glue” showed much less of an advantage.

The explanation these authors offered was based on the idea that ambiguous words have more lexical entries (essentially one for each distinct, known meaning) than unambiguous words. According to Rubenstein and colleagues, the identification of a word involves a search of lexical memory. That search is essentially random, however, it involves only those words that have some orthographic resemblance to the presented word. Because ambiguous words have more lexical entries, the chance of discovering one of these entries early in processing is heightened. As a result, word identification is more rapid.

12.1.2 Clark’s (1973) response

These results did not, of course, go unchallenged. The most influential challenge was offered by Clark (1973). Clark's now well-known argument is that in order to be able to generalize results from experiments contrasting two sets of stimuli (e.g., ambiguous words and unambiguous words) one has to carry out an analysis in which inter-item variability is considered. In Clark's original paper, his recommended analysis was a quasi-F (F') procedure in which the error terms in ANOVAs involved both subject and item variability. When Clark applied this technique to Rubenstein and colleagues' data, their ambiguity advantages became nonsignificant.

Although Clark's (1973) approach has received considerable criticism (Cohen, 1976; Keppel, 1976; Raaijmakers, Schrijnemakers & Gremmen, 1999; Smith, 1976; Wike & Church, 1976), using item variability when analyzing data from language experiments is now common practice. More importantly, Clark's analysis reopened the question of the existence of an ambiguity advantage. Papers published soon thereafter did nothing to remedy the problem. Forster and Bednall (1976) found small and nonsignificant ambiguity effects with both equiprobable and unequiprobable ambiguous words ("balanced" and "unbalanced", respectively). (Categorization of words into the various conditions was again done by subject ratings.) In contrast, Jastrzembski and Stanners (1975) and Jastrzembski (1981), using meaning counts from dictionaries to determine ambiguity, did observe an ambiguity advantage in a series of experiments, even using Clark's very conservative F' procedure.

12.1.3 Gernsbacher's (1984) challenge

The story, however, took another twist when Gernsbacher (1984) argued convincingly that using meaning counts from dictionaries is not a good way to determine the number of meanings a word has. Using *gauge*, *cadet*, and *fudge* as examples, she noted that the former actually has 30 dictionary definitions while the others each have 15. In her informal survey of college professors, she discovered that they could "on average provide only 3 definitions of the word *fudge*, 2 of the word *gauge*, and 1 of the word *cadet*" (p. 272). In general, it's clear that the match between dictionary definitions and what people actually know about a word isn't very good. More importantly for her purposes, she also noted that in none of the previous demonstrations of an ambiguity advantage had the researchers equated their stimuli on "experiential familiarity", a variable that she had demonstrated to be very important in lexical decision tasks. (See

Gernsbacher, 1984, for a discussion of the experiential familiarity concept.) Using stimuli in which rated experiential familiarity was controlled, Gernsbacher found no evidence of an ambiguity advantage.

12.1.4 Firmly establishing the effect

Gernsbacher's (1984) arguments caused Millis and Button (1989) to re-evaluate the issue. Recognizing that people's knowledge about words' meanings is often different than what's contained in dictionaries, they created three different subjective ratings of the number of available meanings. The first involved having subjects record the first meaning that came to mind when reading a word (essentially, Rubenstein *et al.'s*, 1970, procedure). The second involved having subjects record all the meanings that came to mind when reading a word and counting the total number of meanings across all subjects. The third also involved having subjects record all meanings that came to mind when reading a word but what was counted was the average number of meanings generated per subject. Few meaning and multiple meaning words were selected using each of these measures (equating words sets on experiential familiarity). Lexical decision results showed an ambiguity advantage when ambiguity was defined using either the second or third measure and a large (87 ms) but nonsigificant advantage using the first measure (using Clark's (1983) F' procedure).

Although Millis and Button (1989) failed to replicate Gernsbacher's (1984) null effect, their results do underline one of Gernsbacher's main points. Word properties should be measured by determining how those properties are actually represented in subjects' minds. Since Millis and Button's paper most researchers have done exactly that and the existence of the ambiguity advantage is now reasonably well documented. For example, Kellas, Ferraro and Simpson (1988) demonstrated an ambiguity advantage in two lexical decision experiments when familiarity was controlled and ambiguity was measured by using a procedure those authors developed. In this procedure, subjects are asked to rate words as to whether they have no meaning (0), one meaning (1) or more than one meaning (2). Hino and Lupker (1996), using Kellas *et al.'s* procedure to select stimuli, showed an ambiguity advantage for both high and low frequency words in lexical decision tasks and for low frequency words in naming tasks. Borowsky and Masson (1996) also observed an ambiguity advantage in lexical decision when using Kellas *et al.'s* procedure, although they didn't observe an advantage in their naming task. Nonetheless, ambiguity advantages in naming have been reported by Gottlob, Goldinger, Stone and Van Orden

(1999), Hino, Lupker, Sears and Ogawa (1998), Hino, Lupker and Besner (1998), Lichacz, Herdman, Lefevre and Baird (1999) and Rodd (2004).

Rodd's (2004) data provide a likely explanation for the difference between Borowsky and Masson's (1996) results and those reported elsewhere. Rodd showed that the effect size grows as function of the difficulty of the words being named (i.e., the effect existed for exception words but not for easy-to-name regular words). Certainly, Borowsky and Masson's words were short, typically regular and produced mean naming latencies between 494 and 508 ms, suggesting that any ambiguity advantage would have been quite small for those stimuli. This analysis is also consistent with Hino and Lupker's (1996) observation that, in naming, there was no ambiguity advantage for high frequency words.

12.2 Explaining the ambiguity advantage

12.2.1 Models based on multiple lexical units

The first attempt to explain the ambiguity advantage in lexical decision was Rubenstein *et al.'s* (1970) search model. Because Rubenstein *et al*. did not observe an interaction between frequency and ambiguity, they proposed that the two variables affected separate processing stages. Frequency affected the "marking" process, the process by which lexical units were designated for further evaluation. The higher a word's frequency the sooner it gets marked and enters the set of words available for evaluation. The ambiguity effect, in contrast, emerged because, as noted earlier, each meaning of a word has a separate lexical entry. Thus, on average, the random search through the marked entries would locate one of the multiple lexical entries for an ambiguous word more rapidly than the single lexical entry for an unambiguous word.

No computational version of this model was, of course, created and, hence, it's not entirely clear that the model actually does explain Rubenstein *et al.'s* (1970) results. For example, if each meaning of an ambiguous word has a separate lexical entry, the frequency values of those lexical entries would, presumably, represent the frequency with which that particular meaning had been previously activated. These frequency values would be less than the frequency value of a matched unambiguous word because that word's single meaning would have been activated every time the word was processed. As such, it is more likely that an unambiguous word would be entered into the marked set early in processing. If so, the result would be an ambiguity disadvantage rather than an ambiguity advantage.

The second attempt to explain the ambiguity advantage was proposed by Jastrzembski (1981) who did observe an interaction between frequency and ambiguity. Jastrzembski's account was an activation account based on Morton's (1969) logogen model. According to this model, each word in a reader's lexicon is represented by a logogen. In reading, word identification occurs when the activation of its logogen reaches a threshold value. The activation threshold for each logogen is a function of the word's frequency, higher frequency words have lower thresholds and, hence, reach threshold more rapidly. Jastrzembski's idea, like Rubenstein *et al.'s* (1970), was that ambiguous words have multiple entries (i.e., multiple logogens). Hence, the chance of one of them reaching threshold early would be higher than the chance of the single logogen of an unambiguous word reaching threshold early.

Again, no computational version of the model was available and, thus, it's not entirely clear that it does explain Jastrzembski's (1981) results. The problem is the same as with Rubenstein *et al.'s* model. The separate multiple logogens for each meaning of an ambiguous word would, presumably, have threshold values appropriate to the frequency of that particular meaning. As such, the thresholds for all logogens of an ambiguous word would be higher than the single threshold for the logogen of a frequency-matched unambiguous word. If so, it seems unlikely that an ambiguity advantage would emerge.

Forster and Bednall's (1976) results provide an additional problem for these types of accounts. Forster and Bednall asked subjects to decide whether a word had multiple meanings or not. They found that "yes" responses to ambiguous words were faster than "no" responses to unambiguous words presumably due to the fact that exhaustive searches were necessary in order to be able to say "no" to unambiguous words. What they didn't find was the expected difference between ambiguous words with balanced and unbalanced meanings. Presumably, in order to respond "yes" readers must find both meanings (in a search context) or have both meanings activated over threshold (in an activation context). Thus, the frequency of the less probable meaning should determine response latency. Words with unbalanced meanings should, therefore, suffer in contrast to words with balanced meanings. For high frequency words there was no difference between the two ambiguous words types. For low frequency words, the words with unbalanced meanings showed a nonsignificant 37 ms advantage.

12.2.2 Models based on distributed representations

To this point, the models discussed have all been “localist” models. The term “localist” in this context refers to the assumption that a unit in memory represents a full meaning. More recently, localist models have become less popular and there has been an explosion of models based on distributed representations. According to these types of models, meanings are represented as a pattern of activation across a set of meaning units. Similarly, a word’s orthography is represented as pattern of activation across a set of orthographic units and a word’s phonology is represented as a pattern of activation across a set of phonological units. These units are interconnected and, through a learning process, the connections come to be weighted in a way that reflects the appropriate relationships among the units.

When a word is visually presented, a set of orthographic units is initially activated. This activation then spreads to semantic units through the weighted connections. Most importantly, the degree to which the connection weights come to represent those connections is a function of the consistency of the connections. That is, a set of orthographic units that always activates a specific set of semantic units will produce strong connections among those units (i.e., having 1:1 “feedforward” connections from orthography to meaning builds strong connections). Ambiguous words do not facilitate the development of strong connections because ambiguous words, by necessity, activate different semantic units (i.e., meanings) in different situations (i.e., 1:many “feedforward” connections are weaker). As a result, the general expectation derived from a model with this type of architecture is that meaning activation for ambiguous words will be slower and more error prone than that for ambiguous words.

12.2.3 The problem for distributed representation models

Joordens and Besner (1994) provided one of the first investigations of this issue using Masson’s (1991) parallel distributed processing (PDP) model of semantic memory. Their simulation results showed that, typically, this model was unsuccessful in activating a semantic pattern appropriate to any single meaning, instead activating and ultimately settling into a “blend state”, a combination of the semantics from the two meanings. When the simulations were successful (e.g., typically when the two meanings of the ambiguous word had quite different frequencies), there was no ambiguity effect. Joordens and Besner also reported that when Hinton and Shallice’s (1991) model was examined, performance was actually noticeably better for unambiguous words than for ambiguous words (in terms of error scores).

Kawamoto (1993) and Kawamoto, Farrar and Kello (1994) reported similar results using their PDP model. That is, the time (i.e., the number of processing cycles) taken to activate an appropriate semantic code was longer and the settling process was more error prone for ambiguous words than for unambiguous words. Similarly, Borowsky and Masson (1996), using Masson's (1995) model, reported that full semantic activation was slower and more error prone for ambiguous words than for unambiguous words.

12.2.4 The solutions for distributed representation models

In order to explain the ambiguity advantage, both Kawamoto *et al*. (1994) and Borowsky and Masson (1996) assumed that lexical decision making is not based on the time to activate the appropriate semantic codes (see also Besner & Joordens, 1995, Masson & Borowsky, 1995, Piercey & Joordens, 1999, and Rueckl, 1995). Kawamoto *et al*. assumed that lexical decisions were made on the basis of orthographic activity. They also noted that when their semantic system was trained in a particular way (i.e., using the "least mean square" (LMS) error-correction algorithm rather than the Hebbian learning algorithm), the model could account for the ambiguity advantage. The reason was that when using the LMS algorithm, the orthographic-semantic inconsistencies for ambiguous words caused the connections in the orthographic system to become stronger (essentially making up for the weak orthographic-semantic connections). Thus, processing at the orthographic level was better for ambiguous words than for unambiguous words. (A similar argument could be applied to the phonological connections, allowing the model to explain the ambiguity advantage reported by Hino and Lupker (1996), Lichacz *et al*. (1999) and Rodd (2004) in the naming task.)

A slightly different account was offered by Borowsky and Masson (1996). These authors suggested that lexical decision making was based on computing the sum of energy at the orthographic and semantic levels. When that sum reached a criterion value, a positive decision could be made. Indeed, ambiguous words, due to the semantic activation they produce, do reach this criterion faster in their model, allowing it to predict an ambiguity advantage. Note, however, the model does not predict an ambiguity advantage in naming (i.e., phonological units are activated at the same rate for both ambiguous and unambiguous words). This is intentional. As noted, Borowsky and Masson didn't observe an ambiguity advantage in naming. Thus, various parameters of the model (i.e., those reflecting the phonological-semantic linkages) were set so that the model wouldn't produce an effect. In fact, before doing this, the model tended to predict an ambiguity disadvantage in naming. Reducing the weightings on these parameters nullified this disadvantage. It's not impossible that further reductions would allow the model to explain the ambiguity advantage in naming observed by Hino and Lupker (1996), Lichacz *et al*. (1999) and Rodd (2004).

Our work in this area (Hino & Lupker, 1996; Hino, Lupker & Pexman, 2002; Hino, Pexman & Lupker, 2005; Pexman, Hino & Lupker, 2004; Pexman & Lupker, 1999) has produced a third account of the ambiguity advantage, what we call the "feedback account". The framework we've used is a PDP framework, although the principles could also be applied to localist frameworks like Coltheart, Rastle, Perry, Langdon and Ziegler's (2001) dual-route model. The main requirement is that the system be highly interactive. The idea is that once an orthographic representation starts to become active, semantic activation (and phonological activation) follows rapidly. This semantic activation then feeds back to the orthographic level (and forward to the phonological level) to help stabilize the activity there. Because ambiguous words, by definition, have multiple meanings, on average, they would have more substantial semantic representations. Thus, the semantic feedback (and feedforward) activation they provide will be stronger than that from unambiguous words, allowing the orthographic and phonological activation for ambiguous words to stabilize more rapidly. Under the assumption that orthographic activation drives performance in the lexical decision task while phonological activation drives performance in the naming task, the prediction is an ambiguity advantage in both tasks.

Although there is, as yet, no implemented version of the model, the general principle that feedback plays a major role in word recognition (see Balota, Ferraro & Conner, 1991) does have considerable support. For, example, the principle provides a ready explanation of the homophone disadvantage in lexical decision (e.g., Pexman, Lupker & Jared, 2001), the synonym disadvantage in lexical decision (Hino *et al*., 2002; Pecher, 2001) and the lexical decision and naming advantages for words with larger numbers of features (Pexman, Lupker & Hino, 2002). Thus, at present, of all the accounts mentioned above, the feedback account appears to be the most successful.

12.3 A reconsideration of the concept of ambiguity

12.3.1 The meaning-sense distinction

In the above discussion the concept of ambiguity has been regarded as unidimensional. Words have some number of alternative meanings and readers need to pick the right meaning in order to understand the writer's story. In the Linguistics literature, however, the concept of ambiguity is considered to be more complicated (e.g., Caramazza & Grober, 1976; Nunberg, 1979; Tuggy, 1993). In particular, a clear distinction is made between words that are homonyms, that is, words that have multiple unrelated meanings, and words that are polysemous, that is, words with multiple senses based on the same original meaning. Homonyms are essentially accidents of history. *Bank*, the classic example of an ambiguous word, is a homonym. The fact that it means a place to keep your money and the side of a river is a result of two independent contributions to the English language (probably Germanic and Scandinavian, respectively). *Roll*, on the other hand, is polysemous. The fact that it means a list of names, any of various food preparations rolled up for cooking or serving, a flight maneuver, a heavy reverberatory sound, etc. is not an accident. Each of these senses is derived from the core meaning of the word *roll*.

The idea is that because the multiple senses of polysemous words are derived from the same core meaning, the representation of the different senses in semantic memory should be somewhat intertwined (e.g., they will share features). The same is not true for homonyms which should have two (or more) distinct representations in semantic memory. As a result, the two types of ambiguous words may have different processing implications. In particular, from a PDP perspective, polysemous words would be less likely to cause the blend state problem that seems to occur when words with two distinct meanings are processed (Joordens & Besner, 1994). One could even hypothesize that researchers who found an ambiguity advantage may have done so because they tended to use polysemous words as their ambiguous words while the failures to find an ambiguity advantage could be attributed to the use of homonyms.

12.3.2 Processing implications? - Klein and Murphy (2001; 2002)

The question of whether this linguistic distinction is psychologically real is, of course, a crucial one. To many, the distinction between *bank* and *roll* noted above seems rather artificial. Most of the meanings/senses listed for both words seem to represent quite different meanings. In fact, readers are encouraged to reconsider the example at the beginning of this chapter. Is *belt* a homonym or a polysemous word? The answer is provided near the end of this section of the chapter.

One attempt to directly assess the psychological reality of this distinction was provided by Klein and Murphy (2001; 2002). The main task Klein and Murphy (2001) used was a "sensicality judgment task". Word pairs (e.g., *daily paper, yellow lecture*) were presented and subjects had to decide whether the word combination made sense. The key manipulation involved sequential trials in which the second word was repeated (e.g., *daily paper* - *wrapping paper*). On half of these trials, the first word in the two pairs evoked the same sense of the second word (e.g., *daily paper* - *liberal paper*). On the other half, it evoked a different sense (e.g., *wrapping paper* - *liberal paper*). The idea was that if all the senses of a concept are stored together in memory, both *daily paper* and *wrapping paper* should activate the semantic information necessary to process *liberal paper* effectively. If the senses of *paper* were stored separately, however, pairs evoking the same sense would be much more effective "primes". Across a number of experiments, Klein and Murphy (2001) found a large advantage for same sense primes. In fact, the advantages for same sense primes were essentially the same size as the advantage for same meaning primes when homonyms were used (e.g., *commercial bank* - *savings bank* versus *creek bank* - *savings bank*). They concluded "the main empirical result is the finding that different senses have little functional overlap - about the same as the unrelated meanings of homonyms" (p. 277).

Klein and Murphy (2002) reached a similar conclusion based on results in a similarity judgment task. In this task, subjects were asked to judge which of two two-word phrases was most similar to a target phrase (e.g., *daily paper*). One phrase used the same second word as the target phrase but had a first word that evoked a different sense for the second word (e.g., *shredded paper*). The other phrase did not repeat words, however, the phrase was related to the concept in the target phrase either taxonomically (e.g., *evening news*) or thematically (e.g., *smart editor*). Subjects rarely chose second phrases that shared a word with the target phrase (<20% of the time). In fact, they only chose the second phrase which shared the target word slightly more often in the polysemous word condition than in the homonym condition (e.g. target: *national bank*, options: *river bank* and *checking account*). Klein and Murphy concluded that "different senses of a word are probably related but are not generally similar" (p. 566). In essence, what Klein and Murphy's (2001; 2002) research suggests is that the different senses of a word are represented much more distinctly in memory than one might have imagined.

12.3.3 Processing implications? - Azuma and Van Orden (1997)

In spite of Klein and Murphy's (2001; 2002) findings, it seems reasonable that there could be at least some processing differences for polysemous words versus homonyms, or at least for ambiguous words with related versus unrelated meanings. Azuma and Van Orden (1997) appear to have provided the first specific examination of this question as it relates to the ambiguity advantage in the lexical decision task. In their experiments, Azuma and Van Orden factorially manipulated number of meanings and the relatedness of those meanings. Following Gernsbacher's (1984) arguments that the best way to know what's going on in a subject's mind is to ask, Azuma and Van Orden obtained number-of-meaning and relatedness-of-meaning measures by asking subjects to rate words on both dimensions. To determine the number of meanings, they used Millis and Button's (1989) total meanings metric. In this procedure, subjects write down all the meanings that they can think of for each word. Each meaning is compared against a dictionary definition and a count is made of how many of the dictionary meanings are listed by at least one subject. To determine the relatedness of meanings, they selected each word's dominant meaning and asked subjects to rate how strongly it was related to each of the subordinate meanings (on a seven point scale). They then calculated the average of these ratings.

Using standard nonwords (i.e., the type used in virtually every other experiment investigating these issues in the prior literature (e.g., *prane*)) in their lexical decision task, Azuma and Van Orden got a nonsignificant 8 ms number-of-meanings effect and no hint of a relatedness effect. Using pseudohomophones (nonwords that, when pronounced, sound like words (e.g., *brane*)), however, they got a large interaction. The few meanings, low relatedness words had much slower latencies than the other three word types which had equivalent latencies. A second lexical decision experiment using a new set of words and pseudohomophones as nonwords, produced a similar interaction.

Interpreting the slower latency in the few meanings, low relatedness condition in terms of either of the two experimental factors is somewhat difficult. One could interpret this result as implying that when the relatedness of the various meanings is low, having multiple meanings is quite beneficial (the standard ambiguity advantage). Only when there is strong relatedness is there no ambiguity advantage. Alternatively, one could interpret this result as meaning that there is a relatedness effect when the number of meanings is low but not when the number of meanings is high. (In Experiment 1, a multiple regression analysis suggested that the number-of-meanings effect was slightly stronger than the relatedness effect. In Experiment 2, a similar analysis suggested exactly the opposite.) Azuma and Van Orden (1997) chose to interpret the result as showing a relatedness effect but no number-of-meanings effect. That is, they felt that the way they measured relatedness (ratings of how related each subordinate meaning is to the dominant meaning) didn't adequately capture the relatedness among meanings of multiple-meaning words (because the relatedness among subordinate meanings was not considered). Thus, the data from the few-meanings conditions, showing a relatedness advantage, should be taken more seriously than the data from the multiple-meanings condition.

Following Azuma and Van Orden's (1997) logic and only considering the data from the few-meanings conditions creates an obvious problem. There is no way to evaluate the ambiguity (i.e., number-of-meanings) effect. That is, there is no comparison available to determine whether multiple-meaning words are easier to process than few-meaning words. However, the paper does raise two important points. First, relatedness of meanings might be important in lexical decision even if it isn't in Klein and Murphy's (2001; 2002) tasks. Second, the nature of the nonwords used may be important. Indeed, using consonant strings (e.g., *prvnt*) as nonwords inevitably reduces overall latencies and shrinks the size of virtually any effect (Borowsky & Masson, 1996; Stone & Van Orden, 1993). Not surprisingly, when the discrimination is easy there is little time for variables that normally affect processing to show an impact. In contrast, when pseudohomophones are used, latencies are longer and, often, the impact of variables increases. Indeed, Pexman and Lupker (1999) demonstrated this to be the case for the ambiguity advantage.

Azuma and Van Orden (1997) suggested, however, that pseudohomophones do more than simply make the task more difficult. Supposedly, they also get subjects to attend more to semantic information which can, potentially, provide a better window on the nature of semantic representations. Thus, the argument is that the interaction showed up in Azuma and Van Orden's pseudohomophone condition not because the task was more difficult, but because the qualitative nature of the process of distinguishing between words and nonwords changed. Speaking against this argument, however, is the fact that nonsemantic effects also grow in the presence of pseudohomophones. For example, the homophone effect (low frequency homophones like *maid* have longer latencies than nonhomophonic control words) also grows when pseudohomophones are used as nonwords

(Pexman & Lupker, 1999; Pexman *et al*., 2001). Thus, whether pseudohomophones do cause subjects to recruit more semantic information or not, that clearly isn't the only thing they do.

12.3.4 Processing implications? - Rodd, Gaskell and Marslen-Wilson (2002)

Working on the principles postulated by Azuma and Van Orden (1997), Rodd, Gaskell and Marslen-Wilson (2002) took the argument one step further. They postulated that it is only the "senses" of a word that produce the ambiguity advantage. In line with what PDP models typically predict, they further proposed that multiple unrelated meanings actually cause difficulty for activating the appropriate semantic information and, hence, inhibit lexical decision making.

Rodd *et al*. (2002) used the Wordsmyth dictionary, rather than subjective ratings, to determine both how many unrelated meanings and how many senses each of their words had. In their Experiment 2, they factorially manipulated the number of meanings (one or two) and number of senses (few or many - summed over both meanings for two-meaning words) while using pseudohomophones as their nonwords. Rodd *et al*. observed a significant 14 ms number-of-senses advantage and a nonsignificant 6 ms number-of-meanings disadvantage. Based on these results, they claimed that number of senses is the key to the ambiguity advantage and that multiple meanings do cause the types of problem predicted by PDP models. (More recently, Beretta, Fiorentino & Peoppel, 2005, using Rodd *et al.'s* stimuli have reported a number-of-meanings disadvantage (16 ms) that was significant over subjects although not over items.)

In their Experiment 3, Rodd *et al*. used most of the same words in an auditory lexical decision task. In this task, both main effects were now significant and essentially equivalent in size. The most interesting (and novel) aspect of these results is, of course, the significant ambiguity disadvantage. Although there have been failures to replicate the ambiguity advantage in the literature (e.g., Forster & Bednall, 1976; Gernsbacher, 1984), there doesn't seem to be any result even hinting at an ambiguity disadvantage in the lexical decision task. What must be kept in mind, however, is that the auditory lexical decision task is somewhat novel in the word recognition literature. At this point in time, it's less than clear whether the processes involved in making auditory versus visual lexical decisions are similar or not.

This concern emerges more clearly when one reconsiders the issue of the type of nonwords used in these tasks. As noted, Azuma and Van Orden (1997) claimed that their effects (which Rodd *et al*., 2002, argued, are sense effects) only emerge when pseudohomophones are used. Rodd *et al*. appeared to accept Azuma and Van Orden's argument which led to their use of pseudohomophones in their Experiments 1 and 2. In fact, Rodd *et al*. reported in a footnote that when they didn't use pseudohomophones with the words from their Experiment 2, their pattern was even weaker. Pseudohomophones, of course, can't be used in auditory lexical decision tasks. If it sounds like a word, it is a word. Thus, standard nonwords had to be used in Rodd *et al.'s* Experiment 3, which, presumably, should have made it harder, rather than easier, to get their effects. The fact that the ambiguity disadvantage was significant only in Experiment 3 does imply that the visual and auditory lexical decision tasks are based on somewhat different processes.

What should also be noted is that Rodd *et al*. (2002) faced a daunting task in defining their independent variables. First of all, as noted previously, Gernsbacher (1984) provided a rather compelling argument against using dictionary based measures of the number of meanings. The same argument would certainly apply to the count of the number of senses. To return to the earlier question, is *belt* polysemous or a homonym? Are the definitions of *belt* noted earlier separate meanings or different senses? According to Wordsmyth, Rodd *et al.'s* source, *belt*, one of Rodd *et al.'s* words, is polysemous. Those apparently different meanings are actually different senses. It's far from clear that human raters would agree. Equally importantly, defining the "sense" variable in Rodd *et al.'s* manipulation itself is a challenge. Should an ambiguous word with two separate meanings, each with six senses, be thought of as having twelve senses, as Rodd *et al*. assumed? (How about an ambiguous word with twelve separate meanings, each having only one sense?) If so, the matching word in the unambiguous condition should also have twelve senses. However, if one wishes to argue, as Rodd *et al*. do, that what produces the ambiguity advantage is the nature of the representation of the meaning that ultimately is settled on, then that meaning (whichever one it is) only has six senses. Thus, the matching word in the unambiguous condition should only have six senses. Methodologically, it's not clear that there actually is a solution to this problem (although see Jastrzembski, 1981, for a notable attempt).

12.3.5 A model - Rodd, Gaskell and Marslen-Wilson (2004)

Recently, Rodd, Gaskell and Marslen-Wilson (2004) have proposed a PDP model of semantic processing that, with the right assumptions, produces a sense advantage and an ambiguity disadvantage. The idea is that if semantic processing does not have to be completed, as may be the case in the lexical decision task, there is a time period when many sense words would show an advantage due to the nature of their semantic representations while multiple meaning words would show a disadvantage due to the competition created by having separate semantic representations. Thus, lexical decisions made at this point in time would show both a sense advantage and an ambiguity disadvantage. As processing continues, however, even many sense words should start to show a disadvantage due to the fact that the multiple senses ultimately compete with one another as well. Interestingly, the problem of equating words on the total number of senses (discussed in the last paragraph) was handled differently in the model simulation than it was in Rodd *et al.'s* (2002) experiments. In the model simulation, all meanings with multiple senses were assumed to have the same number of senses per meaning. As a result, two-meaning (i.e., ambiguous) multiple sense words had twice as many total senses as unambiguous multiple sense words.

With respect to predictions, the model does produce both a sense advantage and an ambiguity disadvantage. The ambiguity disadvantage, however, was far larger (156 cycles) than the sense advantage (108 cycles) in the simulations Rodd *et al*. (2004) reported. In their Experiment 2, Rodd *et al*. (2002) had observed exactly the opposite relationship while in their Experiment 3, the two effects were nearly the same size. Whether the model can be altered to account for these discrepancies is a matter for future research.

12.4 Ambiguity effects in semantic tasks

12.4.1 Looking for an ambiguity disadvantage

One thing shared by all the models discussed above is that they predict an ambiguity disadvantage in semantically-based tasks (i.e., tasks that require complete, or nearly complete, activation of semantics). That is, although an ambiguity (or sense) advantage might be observed in lexical decision if responding can be accomplished without completing semantic processing, things are different in semantically-based tasks. In localist models, when an ambiguous word is processed, the wrong meaning may be activated, which should hinder performance. In PDP models, ambiguous words produce a competition between meanings, slowing the semantic activation process. Thus, a key question is whether there is an ambiguity disadvantage in such tasks.

There appear to be three types of experimental paradigms that have been used to address this question. The general finding has been, as predicted, an ambiguity disadvantage. The first paradigm involves a standard reading task. Various eye behaviors are monitored as subjects read sentences containing either ambiguous or unambiguous words. At present, consider only situations where the preceding context is purposely neutral so as not to bias the reader toward one meaning of the ambiguous word (e.g., *He thought that the punch/cider* was a little sour).

Using this procedure, Rayner and Duffy (1986) (see also Duffy, Morris & Rayner, 1988, and Rayner & Frazier, 1989) demonstrated that gaze durations on both the target word and the following words were longer when the target word was ambiguous (i.e., *punch* versus *cider*), however, this difference was only observed if the two meanings of the target word were approximately balanced. When one meaning dominated, the difference in target processing disappeared although the difference for post-target words did not. The reason for the post-target difference was that the post-target context was always set up to be supportive of the less dominant meaning of the ambiguous word. The conclusion, therefore, is that there typically is an ambiguity disadvantage when reading for meaning. Only when one meaning is quite dominant (and appropriate) could it, and subsequent text, be read without cost.

The second paradigm is the association (or relatedness) judgment task. In this task, two words (e.g., *bat - vampire*) are presented and subjects must decide whether the words are related. Using sequential presentations with the ambiguous word (e.g., *bat*) as either the first or second presented word, Gottlob *et al*. (1999) demonstrated that it took longer to determine that two words were related when one was ambiguous (see also Piercey & Joordens, 2001). Again, the conclusion is that when the meaning of ambiguous words must be determined, there is a cost.

The third paradigm is the semantic categorization task. Subjects must determine whether each presented word is a member of a designated semantic category. Hino *et al*. (2002) demonstrated that it was more difficult to categorize ambiguous, versus unambiguous, words as being nonliving things in a two-choice (living-nonliving) task. Ambiguity was not manipulated for the living thing stimuli. These results also support the claim that when the meaning of ambiguous words must be determined, there is a cost.

12.4.2 Explaining the ambiguity disadvantage

Unfortunately, answers are never this simple. In all of these tasks, not only must meaning be activated, subjects must also engage decision processes. Thus, in order to determine whether or not having multiple meanings slows the meaning activation process, as both localist and PDP type models would have it, the potential impact of competition during decision-making must be considered. For example, when one fixates on an ambiguous word in text with no prior context to disambiguate it, time and effort would be required to decide which meaning the writer had in mind (unless one meaning is highly dominant, in which case, as noted above, the fixation is short). When an ambiguous word (e.g., *bat*) is presented as the first word in a relatedness judgment task, the selection of the meaning unrelated to the second word (i.e., baseball) would certainly lead to a response delay when the second word (i.e., *vampire*) is presented, as the subject goes back and evaluates alternative meanings of *bat*. When the ambiguous word is presented second, as it was in Gottlob *et al.'s* (1999) Experiment 3, a slightly different problem arises. The activation of two meanings for *bat* early in processing would presumably cause a response conflict. The animal meaning of *bat* suggests that it is related to *vampire*, however, the baseball meaning suggests that a negative response is required. In all cases, an ambiguity disadvantage would emerge, but, not necessarily because of competition during meaning activation, rather because of problems created during decision making.

Indeed, Duffy and Rayner (1986) did propose that the ambiguity disadvantage in gaze duration for unbiased ambiguous words in their experiments may have been due to decision/selection difficulties (see also Frazier & Rayner, 1990). When there is no context, which of the multiple activated meanings do subjects select? With respect to the relatedness judgment task, Pexman *et al*. (2004) have provided clear evidence that the ambiguity disadvantage observed in that task is a response bias effect. Using both sequential and simultaneous presentations of word pairs, Pexman *et al*. replicated Gottlob *et al.'s* (1999) and Piercey and Joordens's (2001) ambiguity disadvantage on positive trials, however, they found no evidence of a disadvantage on negative trials (e.g., *bat* - *door*). The important aspect of negative trials is that ambiguous words create no response conflict on these trials. Both meanings are unrelated to the meaning of the paired word. If the ambiguity disadvantage were due to difficulty activating the meaning for *bat*, there should have been a disadvantage on these trials as well. The implication is that the disadvantage on positive trials is due to response conflict.

The one set of results seemingly immune to this problem is Hino *et al.'s* (2002) demonstration of an ambiguity disadvantage on negative (nonliving) trials in a semantic categorization task. However, as noted, answers are almost never simple. Forster (1999) reported no evidence of any ambiguity effect on negative trials in a different semantic categorization task (i.e., animal-nonanimal). More recently, Hino *et al*. (2005) reexamined this issue and discovered that Forster's results are the more typical results. That is, Hino *et al*. (2005) discovered that there is no ambiguity disadvantage when the task involves small, well-defined categories (e.g., vegetables or animals) and, further, even when using larger, ill-defined categories (e.g., living things), the disadvantage only emerges when considering homonyms (words with unrelated meanings).

If the ambiguity disadvantage were due to difficulty during the meaning activation process, it should show up whenever semantic processing of an ambiguous word is required. The fact that it had such a limited role in Hino *et al.'s* (2004) experiments is more consistent with a decision-making/meaning selection explanation. That is, only when two (or more) completely unrelated meanings become activated in a complicated decision-making process (e.g., *bank* - is it living?) does a delay occur because each meaning has to be thoroughly considered. When the multiple meanings are more closely related or when the categorization task is easier (e.g., animal-nonanimal, vegetable-nonvegetable), a more parallel analysis of the multiple meanings can be done. Thus, no ambiguity disadvantage emerges. The implication, of course, is that the prediction of both localist and PDP models is wrong. There is not an ambiguity disadvantage whenever the semantic activation process must be completed.

This conclusion does produce a rather unfortunate state of affairs. Hopefully, more sophisticated theories will soon emerge to allow for a reasonable explanation of how multiple meanings of a word are represented and activated. One possible basis for such a theory is the work of H. Damasio and colleagues (Damasio, 2001; Damasio, Grabowski, Tranel, Hichwa, & Damasio, 1996; Tranel, Damasio & Damasio, 1997; 1998). This work suggests that different types of semantic information (e.g., nouns versus verbs, different categories of concrete objects) are localized in different neural regions. Activation in each region could certainly arise independently. Thus, when processing a multiple meaning word, the interaction among semantic units that leads to the prediction of a delay in semantic activation may not inevitably exist because the linkages between those units may not exist (i.e., the two meanings are stored quite separately). Obviously, this is also an issue for future research.

12.5 Ambiguity in context

12.5.1 Implications from priming experiments

As noted, a major issue in the psychological investigation of ambiguity has been the impact of context. Can a biasing context alter the meaning activation process for either the biased or unbiased meaning (or are effects of context merely decision-making/meaning-selection effects) and is this process affected by the dominance relationship between meanings? Although a number of experimental paradigms have been used to investigate these issues (e.g., phoneme monitoring (Foss, 1970; Foss & Jenkins, 1973), memory tasks involving rapid serial visual presentations of words (Holms, Arwas & Garrett, 1977), etc.), the most compelling types of experiments are those that tap more closely into on-line processing. Most of these are “priming” experiments in which an ambiguous word is presented in either a neutral or biasing context and is followed by a target that is related to one of the meanings. Subjects are required to make either a lexical decision, naming or color-naming response to the target. We will also consider experiments, like those discussed above, in which eye movements are monitored while people read ambiguous words.

12.5.1.1 Multiple meaning activation independent of context

Possibly the classic paper in this “priming” literature was Conrad’s (1974). Conrad used a modified Stroop (1935) color-naming task. An ambiguous final word in a spoken sentence was followed immediately by a target word written in a color. Subjects were required to name the color. Color-naming latencies were longer for words related to either meaning of the ambiguous word than to matched control words regardless of whether the context was biased toward one meaning or not. Similar results were obtained by Whitney, McKay, Kellas and Emerson (1985). The implication is that all meanings of an ambiguous word appear to be activated when that word is read even if the context is biased toward one of those meanings. As Oden and Spira (1983) demonstrated, however, when there is a delay before the target is presented, context does start to play a role with targets related to the biased meaning producing larger effects.

Swinney (1979) reported similar results using a lexical decision task. Swinney's subjects listened to sentences over headphones while, from time to time, letter strings would appear on the screen for a lexical decision response. The sentence contained an ambiguous word that served as a prime. Immediately following the offset of the ambiguous word, the lexical decision target appeared on the screen. When the sentence was essentially neutral (*The man was not surprised when he found several* <u>*bugs*</u> *in the corner of his room.*) responses to targets related to either meaning of the ambiguous word (e.g., *spy* or *ant*) were faster than responses to unrelated words. Once again, an important finding was that the same result arose even when the sentence was biased (*The man was not surprised when he found several spiders, roaches and other* <u>*bugs*</u> *in the corner of his room*). When the target was delayed three syllables, however, the priming remained for the target appropriate for the context but disappeared for the target that was inappropriate for the context.

Swinney's (1979) (and Conrad's, 1974) basic pattern has now been replicated many times. For example, Blutner and Sommer (1988) provided a direct replication of Swinney's results. Kintsch and Mross (1985), Till, Mross and Kintsch (1988) and Elston-Güttler and Friederici (2005) reported similar results using visual presentations of context sentences. Tanenhaus *et al*. (1979) reported similar results using auditory presentations and a target naming task while Onifer and Swinney (1981) showed that the priming occurred in an immediate target condition even when the sentence was biased toward the dominant meaning and the target was related to the subordinate meaning. (This factor had been uncontrolled in Swinney, 1979.)

12.5.1.2 Evidence that context can affect meaning activation

The implication of all these results is that, initially, all meanings are automatically activated with context playing virtually no role in that process. Again, however, answers are never that simple as contradictory results have also been reported. For example, Simpson (1981) showed that when one used a strong prior context, there was only priming for associates of the intended meaning of the ambiguous word, regardless of whether the targets were associated to the dominant or subordinate meaning. Simpson and Krueger (1991) obtained similar results in a naming task.

Although Simpson's (1981) experiment was criticized because there was a 120 ms lag between the prime word and the target, allowing, in theory, time for context to deactivate the inappropriate meaning, results reported by Tabossi *et al*. (1987), Tabossi (1988) and Tabossi and Zardon (1993) (see also Experiment 2 in Seidenberg *et al*., 1982) don't appear to have this problem. Using essentially the same paradigm as Swinney (1979), Tabossi *et al*. showed that a prior context biased toward the dominant meaning of the ambiguous target only primed associates of the dominant meaning. A parallel type of prior context biased toward the subordinate meaning primed both types of targets. Tabossi further produced data suggesting that getting selective priming of dominant meaning associates required having a context that "makes salient a characteristic feature of it" (p. 334). Other types of context produced essentially equivalent priming of associates of both dominant and subordinate meanings.

Based on Simpson and colleagues' (Simpson, 1981; Simpson & Kreuger, 1991) and Tabossi and colleagues' (Tabossi *et al*., 1987; Tabossi, 1988; Tabossi & Zardon, 1993) results, it appears that, although multiple meanings of ambiguous words are normally activated, it's not impossible for context to suppress activation of the subordinate meaning. However, given the restricted conditions under which such a result has been obtained and the fact that context seems to have no impact on activation of the dominant meaning, a second conclusion would be that the ability of context to influence the meaning activation process is minimal at best.

12.5.2 Measuring eye movements

At first glance, the conclusion that context has an extremely limited role in the meaning activation process, may appear to be at odds with the results in the eye movement literature. In particular, Rayner and colleagues' results (Duffy *et al*., 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989; Rayner, Pacht & Duffy, 1994; Staub & Rayner, this volume) clearly indicate that context does affect fixation times. In Rayner and colleagues' experiments, although balanced ambiguous words appearing in neutral contexts produced longer fixations (compared to control words), a biasing context eliminated this effect. In addition, although ambiguous words with a dominant meanings were read just as rapidly as unambiguous words when the prior context was neutral, when the prior context was biased toward the subordinate meaning there was a cost when reading the ambiguous word. Thus, these results appear to be consistent with the idea that context does affect activation of both dominant and subordinate meanings.

As noted earlier, however, it's unclear whether effects of this sort are due to meaning activation or meaning selection/decision processes. That is, it's possible that the context does affect how rapidly the contextually appropriate meaning is activated. However, it's also possible that the context preceding a balanced ambiguous word merely biases the reader's decision process toward the selection of the intended meaning, eliminating the ambiguity disadvantage. As well, it's possible that, when viewing an unbalanced ambiguous word, a context biased toward the subordinate meaning makes it difficult to ignore that meaning initially, creating decision problems. Thus, while Rayner and colleagues' results certainly speak to the complexity of the reader's ambiguity resolution process, at present, it isn't possible to rule out either an activation-based or a decision-based explanation of those results.

12.5.3 Finding a resolution

In deciding how to determine which explanation is the best explanation of the impact of context, a couple of points should be considered. First, the priming experiments with the shortest SOAs would seem to provide the best window on the activation process because they have the best chance of tapping into the process before a decision has been made (i.e., while all meanings might still be active). Second, as noted, the far more typical finding in those experiments is that, even in biased contexts, both meanings of ambiguous words are activated (e.g., Blutner & Sommer, 1988; Conrad, 1974; Elston-Güttler & Friederici, 2005; Kintsch & Mross, 1985; Onifer & Swinney, 1981; Swinney, 1979; Tanenhaus *et al*., 1979; Till, Mross & Kintsch; 1988). Third, as Hino, Lupker and Sears (1997) have noted, in those priming experiments in which there was little evidence of activation for words related to the subordinate meaning (e.g., Simpson 1981: Simpson & Burgess, 1985), typically no effort was made to equate the strengths of association between the ambiguous word primes and the targets that were related to the dominant versus subordinate meanings. When Hino *et al*. did make such an effort in a single word priming experiment, they showed equivalent priming for the two types of visual targets at a 0 ms interstimulus interval (using auditory primes in a lexical decision task). Thus, it's certainly possible that even in the experiments in which it did appear that the context suppressed activation of the subordinate meaning, it's possible that the effect was simply due to using stronger associates in the dominant meaning condition. Everything considered, it seems more likely that the effects in the eye movement literature are due to decision/selection processes rather than to context affecting the meaning activation process.

12.6 Further thoughts on the activation-selection distinction

Using the eye movement paradigm, Pickering and Frisson (2001) recently reported no cost on initial gaze durations for ambiguous words having two verb meanings (in contrast to Rayner & Duffy's, 1986, results using ambiguous words with two noun meanings). The cost showed up later in the sentence. Pickering and Frisson explained these results by suggesting that because verbs are harder to interpret than nouns and their meanings are often dependent on subsequent words in the sentence, readers will delay meaning selection until they have seen some of those other words.

The noun-verb differences that Pickering and Frisson (2001) reported (see also Seidenberg *et al*., 1982, and Folk & Morris, 2003) obviously represent an interesting avenue for future research. More importantly, because these results suggest that meaning selection can be delayed when it's useful to do so, they underline the point that theories of ambiguity resolution need to distinguish between activation and selection processes. Many of the early researchers in the field, virtually all of whom were working with localist frameworks, were careful to make the activation-selection distinction. For example, the nonselective priming at short SOAs discussed above was typically taken as being informative about the meaning activation process while the selective priming at longer SOAs was taken as being informative about the meaning selection process.

To implement this activation-selection distinction within a localist framework, one has to assume that multiple lexical units are activated initially and they then maintain their activation (and, hence, their candidacy) during the selection process. One would also need to assume that intra-lexical inhibition processes are not so strong that they cause competing candidates to be inhibited. For PDP models (e.g., Rodd *et al*., 2004), the situation appears to be slightly more complicated. For these types of models, the activation process is, inherently, a process of deactivating competitors. Once the activation process runs to completion, the set of semantic units that have been activated defines the meaning that must be selected. Only if the system ends up in a "blend state" (a situation that PDP modelers have assumed represents a failure of the model) are there competitors to select among. Thus, PDP models may have some difficulty with the idea that multiple meanings are maintained for a period of time awaiting the context that allows an accurate selection. In the end, however, it seems clear that what can be thought of as "selection processes" must play a major role in the ambiguity resolution process. Thus, any successful model of that process will need to explain not only how multiple meanings come to be activated but also how readers use context to (usually) successfully select the meaning the writer intended.

12.7 Final thought

Semantic ambiguity is a fact of life for readers/speakers of most languages. The fact that our processing systems seem to allow resolution of these ambiguities so rapidly that we hardly notice them is testimony to a very sophisticated set of language skills. Any successful model of the processes will have to be rather sophisticated as well. The challenge of explaining the ambiguity advantage in lexical decision and naming has certainly provided a large impetus in the development of those models. The challenge of explaining the impact of context provides an additional impetus and should continue to do so. The path ahead has many theoretical twists and methodological turns and successes will be measured in small increments. Nevertheless, the puzzle of understanding how readers deal with ambiguity is one that researchers will continue to find irresistible.

Acknowledgements

Much of the author's research discussed in this chapter was supported by Natural Sciences and Engineering Research Council of Canada Grant A6333. I would like to thank Yasushi Hino, Sachiko Kinoshita and Penny Pexman for their contributions to that research and for their helpful comments on earlier versions of this chapter.

References

Azuma T and Van Orden GC (1997). Why SAFE is better than FAST: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language*, **36**, 484-504.

Balota DA, Ferraro RF, and Connor LT (1991). On the early influence of meaning in word recognition: A review of the literature. In PJ Schwanenflugel ed. *The psychology of word meanings* (pp. 187-221). Hillsdale, N. J.: Lawrence Erlbaum Associates.

Beretta, A, Fiorentina, R and Poeppel, D (2005). The effects of homonymy and polysemy on lexical access: an MEG study. *Cognitive Brain Research*, ***24***, 57-65.

Besner D and Joordens S (1995). Wrestling with ambiguity – Further reflections: Reply to Masson and Borowsky (1995) and Rueckl (1995). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 515-519.

Blutner R and Sommer R (1988). Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language*, **27**, 359-367.

Borowsky R and Masson MEJ (1996). Semantic ambiguity effects in word identification. *Journal of Experiment Psychology: Learning, Memory, and Cognition*, **22**, 63-85.

Caramazza A and Grober E (1976). Polysemy and the structure of the subjective lexicon. In C. Rameh ed., *Georgetown University roundtable on languages and linguistics. Semantics: Theory and application* (pp.181-206). Washington, DC: Georgetown Univ. Press.

Clark HH (1973). The language-as fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, **12**, 335-359.

Cohen J (1976). Random means random. *Journal of Verbal Learning and Verbal Behavior*, **15**, 261-262.

Collins AM and Loftus EF (1975). A spreading activation theory of semantic processing. *Psychological Review*, **82**, 407-428.

Coltheart M, Rastle K, Perry C, Langdon R and Ziegler J (2001). DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, **108**, 204-256.

Conrad C (1974). Context effects in sentence comprehension: A study of the subjective lexicon. *Memory & Cognition*, **2**, 130-138.

Damasio H (2001). Words and concepts in the brain. In J Branquinho ed., *The foundations of cognitive science* (pp. 109-120). Oxford, UK: Oxford University Press.

Damasio H, Grabowski TJ, Tranel D, Hichwa RD and Damasio A (1996). A neural basis for lexical retrieval. *Nature*, **380**, 499-505.

Duffy SA, Morris RK and Rayner K (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language*, **27**, 429-446.

Elston-Güttler KE and Friederici AD (2005). Native and L2 processing of homonyms in sentential context. *Journal of Memory and Language*, **52**, 256-283.

Folk JR and Morris RK (2003). Effects of syntactic category assignment on lexical ambiguity resolution in reading: An eye-movement analysis, *Memory & Cognition*, **31**, 87-99.

Forster KI (1999, November). *Beyond lexical decision: Lexical access in categorization tasks*. Paper presented at the 40th Annual Meeting of the Psychonomic Society, Los Angeles, CA.

Forster KI and Bednall ES (1976). Terminating and exhaustive search in lexical access. *Memory & Cognition*, **4**, 53-61.

Foss DJ (1970). Some effects of ambiguity upon sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, **9**, 699-706.

Foss DJ and Jenkins CM (1973). Some effects of context on the comprehension of ambiguous sentences. *Journal of Verbal Learning and Verbal Behavior*, **12**, 577-589.

Frazier L and Rayner K (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, **29**, 181-200.

Gernsbacher MA (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, **113**, 256-281.

Glucksberg S, Kreuz FJ and Rho SH (1986). Context can constrain lexical access: Implications for models of language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **12**, 323-335.

Gottlob LR, Goldinger SD, Stone GO and Van Orden GC (1999). Reading homographs: Orthographic, phonologic, and semantic dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, **25**, 561-574.

Hino Y and Lupker SJ (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance*, **22**, 1331-1356.

Hino Y, Lupker SJ and Besner D (1998, November). *Polysemy effects in the naming of Japanese Katakana words and their Hiragana transcriptions.* Presented at the 39th Annual Meeting of the Psychonomic Society, Dallas, TX, U.S.A.

Hino Y, Lupker SJ and Pexman PM (2001, November). *Effects of polysemy and relatedness of meanings in lexical decision and semantic categorization tasks*. Presented at the forty-second Annual Meeting of the Psychonomic Society, Orlando, Florida.

Hino Y, Lupker SJ and Pexman PM (2002). Ambiguity and synonymy effects in lexical decision, naming and semantic categorization tasks: Interactions between orthography, phonology and semantics. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **28**, 686-713.

Hino Y, Lupker SJ and Sears CR (1997). The effects of word association and meaning frequency in a cross-modal lexical decision task: is the priming due to "semantic" association? *Canadian Journal of Experimental Psychology*, **51**, 195-210.

Hino Y, Lupker SJ, Sears CR and Ogawa T (1998). The effects of polysemy for Japanese katakana words. *Reading and Writing: An Interdisciplinary Journal*, **10**, 395-424.

Hino Y, Pexman PM and Lupker SJ (2005, submitted). Effects of ambiguity and relatedness of meanings in lexical decision and semantic categorization tasks: Is the speed of semantic coding modulated by the nature of orthographic-to-semantic mappings?

Hinton GE and Shallice T (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, **98**, 74-95.

Hogaboam TW and Perfetti CA (1975). Lexical ambiguity and sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, **14**, 265-274.

Holmes VM, Arwas R and Garrett MF (1977). Prior context and the perception of lexically ambiguous sentences. *Memory & Cognition*, **5**, 103-110.

Jastrzembski JE (1981). Multiple meanings, number of related meanings, frequency of occurrence, and the lexicon. *Cognitive Psychology*, **13**, 278-305.

Jastrzembski JE and Stanners RF (1975). Multiple word meanings and lexical search speed. *Journal of Verbal Learning and Verbal Behavior*, **14**, 534-537.

Joordens S and Besner D (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **20**, 1051-1062.

Kawamoto AH (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language*, **32**, 474-516.

Kawamoto AH, Farrar WT and Kello C (1994). When two meanings are better than one: Modeling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 1233-1247.

Kellas G, Ferraro FR and Simpson GB (1988). Lexical ambiguity and the timecourse of attentional allocation in word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, **14**, 601-609.

Keppel G (1976). Words as random variables. *Journal of Verbal Learning and Verbal Behavior*, **15**, 263-265.

Kintsch W and Mross EF (1985). Context effects in word identification. *Memory & Cognition*, **24**, 336-349.

Klein DE and Murphy GL (2001). The representation of polysemous words. *Journal of Memory and Language*, **45**, 259-282.

Klein DE and Murphy GL (2002). Paper has been my ruin: conceptual relations of polysemous senses. *Journal of Memory and Language*, **47**, 548-570.

Lichacz FM, Herdman CM, Lefevre J-A and Baird B (1999). Polysemy effects in word naming. *Canadian Journal of Experimental Psychology*, **53**, 189-193.

Masson, MEJ (1991). A distributed memory model of context effects in word identification. In D Besner and G Humphreys eds. *Basic processes in reading: Visual word recognition* (pp. 233-263). Hillsdale, NJ: Erlbaum.

Masson MEJ (1995). A distributed memory model of semantic priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 3-23.

Masson MEJ and Borowsky R (1995). Unsettling comments about semantic ambiguity in connectionist models: Comments on Joordens and

Besner (1994). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 509-514.

Millis ML and Button SB (1989). The effect of polysemy on lexical decision time: Now you see it, now you don't. *Memory & Cognition*, **17**, 141-147.

Morton J (1969). Interaction of information in word recognition. *Psychological Review*, **76**, 165-178.

Nunberg G (1979). The non-uniqueness of semantic solutions: Polysemy. *Linguistics and Philosophy*, **3**, 143-184.

Oden GC and Spira JL (1983). Influence of context on the activation and selection of ambiguous word senses. *Quarterly Journal of Experimental Psychology*, **35**, 51-64.

Onifer W and Swinney DA (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency of meaning and contextual bias. *Memory & Cognition*, **15**, 225-236.

Pecher D (2001). Perception is a two-way junction: Feedback semantics in word recognition. *Psychonomic Bulletin & Review*, **8**, 545-551.

Pexman PM, Hino Y and Lupker SJ (2004). Semantic ambiguity and the process of generating meaning from print. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **30**, 1252-1270.

Pexman PM and Lupker SJ (1999). Ambiguity and visual word recognition: Can feedback explain both homophone and polysemy effects? *Canadian Journal of Experimental Psychology*, **53**, 323-334.

Pexman PM, Lupker SJ and Hino Y (2002). The impact of feedback semantics in visual word recognition: Number of features effects in lexical decision and naming tasks. *Psychonomic Bulletin & Review*, **9**, 542-549.

Pexman PM, Lupker SJ and Jared D (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **27**, 139-156.

Pickering MJ and Frisson S (2001). Processing ambiguous verbs: Evidence from eye movements. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **27**, 556-573.

Piercey CD and Joordens S (2000). Turning an advantage into a disadvantage: Ambiguity effects in lexical decision versus reading tasks. *Memory & Cognition*, **28**, 657-666.

Raaijmakers JGW, Schrijnemakers JMC and Gremmen F (1999). How to deal with “the language-as-fixed-effect fallacy”: Common misconceptions and alternative solutions. *Journal of Memory and Language*, **41**, 416-426.

Rastle, K (this volume). Visual word recognition. In GA Gaskell, G Altman, P Bloom, A Caramazza and P Levelt eds. *The Oxford handbook of psycholinguistics*. Oxford, UK: Oxford University Press.

Rayner K, and Duffy SA (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, **14**, 191-201.

Rayner K and Frazier L (1989). Selection mechanisms in reading lexically ambiguous words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **15**, 779-790.

Rayner K, Pacht JM and Duffy SA (1994). Effects of prior encounter and global discourse bias on the processing of lexically ambiguous words. *Journal of Memory and Language*, **33**, 527-544.

Rodd JM (2004). The effect of semantic ambiguity on reading aloud: A twist in the tale. *Psychonomic Bulletin & Review*, **11**, 440-445.

Rodd JM, Gaskell G and Marslen-Wilson WD (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, **46**, 245-266.

Rodd JM, Gaskell MG and Marslen-Wilson WD (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science*, **28**, 89-104.

Rubenstein H, Garfield L and Millikan JA (1970). Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, **9**, 487-49.

Rubenstein H, Lewis SS and Rubenstein MA (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior*, **10**, 645-657.

Rueckl JG (1995). Ambiguity and connectionist networks: Still settling into a solution – Commentary on Joordens and Besner (1994). *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **21**, 501-508.

Seidenberg MS, Tanenhaus MK, Leiman JM and Langer P (1984). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology*, **14**, 489-537.

Simpson GB (1981). Meaning dominance and sementic context in the processing of lexical ambiguity. *Journal of Verbal Learning and Verbal Behavior*, **20**, 120-136.

Simpson GB and Burgess C (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance*, **11**, 28-39.

Simpson GB and Krueger MA (1991). Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, **30**, 627-643.

Smith JEK (1976). The assuming-will-make-it-so fallacy. *Journal of Verbal Learning and Verbal Behavior*, **15**, 262-263.

Staub A and Rayner K (this volume) Eye movements and on-line comprehension processes. In GA Gaskell, G Altman, P Bloom, A Caramazza and P Levelt eds. *The Oxford handbook of psycholinguistics*. Oxford, UK: Oxford University Press.

Stone GO and Van Orden GC (1994). Building a resonance framework for words recognition using design and system principles. *Journal of Experimental Psychology: Human Perception and Performance*, **20**, 1248-1268.

Stroop JR (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, **18**, 643-662.

Swinney DA (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*, **18**, 645-659.

Tabossi P (1988). Accessing lexical ambiguity in different types of sentential context. *Journal of Memory and Language*, **27**, 324-340.

Tabossi P, Colombo L and Job R (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research*, **49**, 161-167.

Tabossi P and Zardon F (1993). Processing ambiguous words in context. *Journal of Memory and Language*, **32**, 359-372.

Tanenhaus MK, Leiman JM and Seidenberg MS (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior*, **18**, 427-440.

Till RE, Mross EF and Kintsch W (1988). Time course of priming for associate and inference words in a discourse context. *Memory & Cognition*, **16**, 283-298.

Tranel D, Damasio H and Damasio A (1997). A neural basis for the retrieval of conceptual knowledge. *Neuropsychologia*, **35**, 1319-1327.

Tranel D, Damasio H and Damasio A (1998). The neural basis of lexical retrieval. In RW Parks, DS Levine and DL Long eds., *Fundamentals of neural network modeling: Neuropsychology and cognitive neuroscience* (pp. 271-296). Cambridge, MA: The MIT Press.

Tuggy D (1993). Ambiguity, polysemy, and vagueness. *Cognitive Linguistics*, **4**, 273-290.

Whitney P, McKay T, Kellas G and Emerson WA Jr. (1985). Semantic activation of noun concepts in context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, **11**, 126-135.

Wike EL and Church JD (1976). Comments on Clark's "The Language-as-Fixed-Effect Fallacy." *Journal of Verbal Learning and Verbal Behavior*, **15**, 249-255.