Semantic Effects in Auditory Word Recognition

Lee H. Wurm*, Douglas A. Vakoch*, Sean R. Seaman*, & Lori Buchanan*

*Wayne State University
*University of California, Davis and SETI Institute
*University of Windsor

Abstract

Until recently most models of word recognition have assumed that semantic effects come into play only after the identification of the word in question. What little evidence exists for early semantic effects in word recognition has relied primarily on priming manipulations, and has used visual stimulus presentation. The current study uses both visual and auditory stimulus presentation, and does not use priming. Response latencies for 100 common nouns were found to depend on the rated Danger and Usefulness of the words. In addition, the interaction between Danger and Usefulness was significant. All effects were above and beyond the effects of concreteness, word length, frequency, onset phoneme characteristics, stress, and neighborhood density. Results are discussed against evidence from several areas of research suggesting a role of behaviorally important information in perception.

1 Introduction

When an organism gathers information about its environment, two characteristics of the objects in that environment are particularly salient for the continued survival of the organism. One is the threat that the object poses. The other, less often mentioned but no less important, is the desirability (i.e., usefulness or value) of the object. Information exchange is beneficial to the extent that the observer learns about both of these characteristics (Darwin, 1859/1968; Davidson, 1992; Schneirla, 1965). The duality of threat and desirability is consistent with evidence that there are two different kinds of reactions to affective stimuli: appetitive reactions and defensive reactions (Bradley, Cossio, Cuthbert, & Lang, 2001; Lang, Bradley, & Cuthbert, 1998). Such research has examined physiological responses (e.g., skin conductance, heart rate, and modulation of the startle reflex) as well as
subjective participant ratings, to visually-presented materials with different kinds of affective content.

In previous work we have explored the possibility that appetitive and defensive responses to stimuli may correspond to aspects of meaning that are coded in lexical entries (Wurm & Vakoch, 1996; Vakoch & Wurm, 1997; Wurm, Vakoch, & Seaman, in press). In some of our earlier work, we have found that lexical decision and naming times are related to the Danger and Usefulness of word referents (Wurm & Vakoch, 2000; Wurm, Vakoch, Aycock, & Childers, 2003). The results of these studies were interpreted in terms of the familiar “approach/avoid” response tendencies.

The question of whether semantic information can influence low-level recognition processes bears on the issue of architectural modularity. Most models of word recognition have not directly incorporated semantics into their architectures, but have postponed the activation and integration of semantic information until some point following the unique identification of a single word. Nevertheless, there is some compelling experimental evidence that semantic information is activated before this time, and some models reflect this. In Marslen-Wilson’s (1987) cohort model, for example, the first 150 msec of speech that a listener hears (a stretch of speech usually not sufficient for word identification) is enough to produce preliminary activation of limited aspects of semantic information for the subset of words consistent with this acoustic input (i.e., for the "cohort"). Zwislocki (1989) agreed with this conclusion based on a series of cross-modal priming experiments in Dutch. In these experiments, an auditory prime word preceded a visual probe word, and participants were required to make a lexical decision about the probe word. She found that presenting the first three phonemes ([kaep]) of kapitein ("captain") led to faster recognition times for visually-presented probe words related semantically to both kapitein and kapitaal ("capital"). This multiple activation occurred even when the sentence context was strongly biased toward one or the other word.

More recently, Tyler, Moss, Galpin, and Voice (2002; see also Moss, McCormick, & Tyler, 1997; Strain, Patterson, & Seidenberg, 1995; Tyler, Voice, & Moss, 2000) studied the effects of concreteness (which the authors consider to be the same variable as imageability in cross-modal priming experiments. Concrete words showed larger priming effects than abstract ones, although both kinds showed significant priming. Importantly, these effects were present at several different probe positions, including one at the identification point (located empirically with gating data). This again suggests very early activation of some aspects of semantic information. The authors suggest that a model like that of Gaskell and Marslen-Wilson (1997) may be able to explain the data. In this model the speech signal is mapped directly
onto a distributed representation that carries all information associated with a word (i.e., both semantics and phonology, along with other things). The connection weights in such a model must encode information about both semantics and phonology, and therefore semantic information will be intricately involved in the recognition process from the start.

The current study extends existing work in this area and differs from it in several ways. First, the present study is not a priming study. We are interested in determining whether the time-course of word recognition is affected by semantic factors, without any priming or context manipulation. Finding such effects would allow a more basic conclusion to be drawn about the organization of lexical memory because it would not be dependent on the possible relationships between pairs of closely related words. Second, previous studies have focused on variables such as concreteness/imageability. In the current study, variables such as these will be controlled for, and a different set of semantic dimensions will be examined. These dimensions carry behaviorally significant information and provide a coherent framework for clarifying subtle nuances of connotative meaning.

Finally, in the current study we will use both auditory and visual stimulus presentation. This has rarely been done in the existing literature, but it is crucial if we want our explanatory frameworks to apply to word recognition in a general way. Nouns will be presented auditorily (Experiment 1) or visually (Experiment 2) to participants who will be required to make a speeded lexical decision about each stimulus.

2 Preliminary rating experiment

Wurm and Vakoch (2000) did a preliminary rating study in which they gathered many of the stimulus values that will be used as regressors in the current study. In that study ratings were obtained from participants for the stimulus words on the semantic dimensions of interest, and values for each stimulus on a number of potential control variables were computed.

2.1 Materials and method

Wurm and Vakoch (2000) asked 64 native speakers of English to rate 100 common nouns on two 8-point semantic scales. Inspection of the Appendix shows that these are common words, likely to be familiar to undergraduate students. One rating scale had endpoints labeled "Not at all Dangerous to Human Survival" and "Extremely Dangerous to Human Survival," and the
other scale had endpoints labeled "Not at all Useful for Human Survival" and "Extremely Useful for Human Survival." Participants were allowed to take as long as they wished to provide their ratings. For each word, the mean Danger rating and the mean Usefulness rating were calculated. The authors also created 100 pseudowords by changing the phoneme at the UP to a different phoneme from the same broad class. For example, the word "poverty" was changed to the pseudoword "pozerty," and "lavender" was changed to "lavelder." Each stimulus was digitized at 20 kHz, low-pass filtered at 9.8 kHz, and stored in a disc file.

2.2 Results and discussion

The mean rating on Danger for the 100 words was 3.47 (SEM = .22), and on Usefulness it was 3.52 (SEM = .18). The two dimensions were moderately but significantly correlated with each other (r = -.25, p < .05). Items covered almost the entire range from 1 to 8 on both dimensions, and the distributions of ratings were similar.

3 Calculation of potential control variables

Following Wurm and Vakoch (2000), we assessed for effects of seven classes of control variables prior to determining whether Danger and Usefulness ratings were related to lexical decision times.

1. Word frequency was computed from CELEX (Baayen, Piepenbrock, & Gullikers, 1995), and familiarity ratings were taken from the MRC Psycholinguistic Database.

2. Word length was computed two ways for each word, in milliseconds and in number of phonemes. Longer words require more time to recognize.

3. Bigram frequency was calculated for each word using the tables of Mayzner and Tresselt (1965).

4. First-syllable stress was coded as either strong (1) or weak (0).

5. Four different measures of neighborhood density were computed. One was an auditory analogue of Coltheart's N (Coltheart, Davelaar, Jonasson, & Besner, 1977), which is the number of words that can be made from a given target word by the substitution of one letter, preserving letter position. For example, if the target word is pin, then pen, win, and pit (along with many others) would be included in the list of neighbors.
The second approach is like Coltheart's N, but also allows for the addition or deletion of one phoneme at the beginning or end of a word. Using the above example, we would also add to the list of neighbors words such as spin and pins.

For both of these measures, the summed frequency of all of a word's neighbors was also computed. This gives a second, frequency-weighted version of each density measure.

6. Concreteness, imagery, and meaningfulness were taken from the MRC Psycholinguistic Database. This database includes two different measures of meaningfulness, both of which were used in the preliminary analyses.

7. Animate or inanimate referent was also included.

These seven classes of variables contain a total of 14 variables. These variables were included as factors in the regression model, except in the cases where there were multiple measures of the same construct. In such cases (e.g., for neighborhood density), the measure that explained the most variance was retained in the analysis. Possible effects of Danger and Usefulness were assessed only after statistically controlling for the effects of these variables.

4 Experiment 1

4.1 Materials and method

Experiment 1 is essentially a replication of Wurm and Vakoch (2000). The 200 stimuli described above were presented in a random order to groups of one to three participants. An additional 230 words were selected for use as filler items, to be included so that the stimulus list would not be comprised entirely of nouns. Half of these items were changed to pseudowords by the same method described above. Filler words were common, recognizable words of various parts of speech, and were from one to five syllables long, as were the critical items. Participants were native speakers of English (students at Wayne State University) who reported normal hearing. Participants were instructed to make a speeded lexical decision, pressing one button on a response box for real words and a different button for pseudowords. A total of 60 students participated in Experiment 1. None of these had participated in the preliminary rating study.

Prior to the main experiment, a practice list of 24 items was presented, in order to familiarize participants with the procedure.
4.2 Results and discussion

Data were discarded for trials on which the word/pseudoword decision was made incorrectly (7.4% of the data). One item ("canteen") was excluded because it was an extreme outlier in the uniqueness point (UP) distribution ($z = 3.7$, $p = .0001$; see Tabachnick & Fidell, 2001). The dependent variable in all analyses was RT on correct trials, for the 100 critical words.

Lexical decision times were measured from the UP of each word. The first author and a research assistant made independent measurements of each UP, which was defined as the middle of the prototypical segment of the particular phone in question (following Radeau, Mously, & Bertelson, 1989; see also Wurm & Ross, 2001). This point was located using both visual and auditory criteria, with the help of a commercial waveform editor. Measurements for a given word were generally within a few msec of each other, and the value used for the UP of each stimulus was the mean of these two independent measurements. UP location was not correlated with Danger or Usefulness ratings.

A hierarchical multiple regression analysis was conducted in which the main effects of Danger and Usefulness were assessed, along with their interaction. Variables were entered in four steps, with simultaneous entry for all variables within a given step. This analytic procedure parallels that of our earlier studies (see also Cohen & Cohen, 1983; Lorch & Myers, 1990).

The results of this analysis are shown in Table 1. At Step 1 of the analysis the between-subjects variance was partitioned. In repeated-measures regression analyses this is done by entering a block of $N - 1$ (i.e., 59 for this experiment) dummy variables that represent the participants. At Step 2 of the analysis the effects of the control variables were statistically removed. This step shows a replication of the concreteness effect reported by Tyler et al. (2000) for auditorily-presented stimuli. In addition, faster lexical decision times were found for words with animate referents, words from sparser lexical neighborhoods, words with lower bigram frequencies, and words with higher word frequencies.

Steps 3 and 4 of the analysis contained the statistical tests of primary interest. At Step 3 the main effects of Danger and Usefulness were assessed. Both were significant. Higher dimension weights on either variable were associated with faster lexical decision times. It should be noted that these main effects are significant over and above the effects of the control variables, including concreteness. Thus, these effects are not due to the semantic variables being confounded with anything at Step 2 of the analysis. Similarly, the interaction (discussed next) was evaluated only after the main effects of the semantic dimensions were controlled for.
Table 1. Summary of Hierarchical Regression Analysis for Variables Predicting Auditory Lexical Decision Time

<table>
<thead>
<tr>
<th>Variable(s)</th>
<th>df</th>
<th>B (SE B)</th>
<th>β</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Between Subjects</td>
<td>59</td>
<td>---</td>
<td>---</td>
<td>18.54***</td>
</tr>
<tr>
<td>2. Control Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concreteness</td>
<td>1</td>
<td>-0.1 (0.01)</td>
<td>-.10</td>
<td>63.63***</td>
</tr>
<tr>
<td>Animate/inanimate</td>
<td>1</td>
<td>-23.8 (5.11)</td>
<td>-.06</td>
<td>21.72***</td>
</tr>
<tr>
<td>Neighborhood density</td>
<td>1</td>
<td>89.6 (6.76)</td>
<td>.20</td>
<td>175.62***</td>
</tr>
<tr>
<td>Item duration</td>
<td>1</td>
<td>7.3 (17.23)</td>
<td>.01</td>
<td>&lt; 1.0</td>
</tr>
<tr>
<td>Bigram frequency</td>
<td>1</td>
<td>17.3 (1.82)</td>
<td>.11</td>
<td>90.16***</td>
</tr>
<tr>
<td>Stress</td>
<td>1</td>
<td>-5.8 (6.41)</td>
<td>-.01</td>
<td>&lt; 1.0</td>
</tr>
<tr>
<td>Word frequency</td>
<td>1</td>
<td>-43.2 (3.36)</td>
<td>-.16</td>
<td>165.12***</td>
</tr>
<tr>
<td>3. Semantic dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danger</td>
<td>1</td>
<td>-1.9 (0.92)</td>
<td>-.02</td>
<td>4.32*</td>
</tr>
<tr>
<td>Usefulness</td>
<td>1</td>
<td>-12.6 (1.22)</td>
<td>-.13</td>
<td>106.58***</td>
</tr>
<tr>
<td>4. Danger x Usefulness</td>
<td>1</td>
<td>1.9 (0.51)</td>
<td>.09</td>
<td>13.85***</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01. ***p < .001.

Figure 1. The slope of the relationship between lexical decision time and Danger, as a function of Usefulness. "Low" indicates a value one standard
deviation below the mean and "High" indicates a value one standard deviation above the mean. The dimensions are continuous; Low and High values were used only for graphical purposes.

The Danger x Usefulness interaction assessed at Step 4 of the analysis was significant, and is shown in Figure 1. The figure was constructed by plotting the regression equation. The mean value of each variable in the statistical model was multiplied by its regression coefficient, except for Danger and Usefulness. For these two variables, the mean plus one standard deviation was used for "High," and the mean minus one standard deviation was used for "Low." Readers are reminded that the underlying analyses were continuous; this is merely the most convenient way to show the nature of the interaction.

The significant interaction indicates that the slope of the relationship between Danger and RT depends significantly on Usefulness. Inspection of Figure 1 shows this dependence. For words rated relatively low on Usefulness, higher levels of Danger are associated with faster RTs. For words rated relatively high on Usefulness, though, higher levels of Danger are associated with slower RTs.

5 Experiment 2

Experiment 2 was conducted in order to determine whether the observed semantic effects are restricted to auditory word recognition.

5.1 Materials and method

Experiment 2 used the 100 common nouns and 100 legal pseudowords described in the Preliminary Rating Study. The same procedure was used as in Experiment 1, but the stimuli were presented visually rather than auditorily. Participants were native speakers of English (students at the University of Windsor) who reported normal vision. A total of 31 students participated in Experiment 2. None of these had participated in the preliminary rating study or in Experiment 1. Prior to the main experiment, a practice list of 24 items was presented, in order to familiarize participants with the procedure.

5.2 Results and discussion

Lexical decision times were measured from the visual onset of each stimulus.
Data were discarded for trials on which the word/pseudoword decision was made incorrectly (7.0% of the data). RTs longer than 1900 msec were also discarded (3.7% of the data). The dependent variable in all analyses was RT on correct trials, for the 100 critical words.

A hierarchical multiple regression analysis analogous to the one reported above was conducted. The results of this analysis are shown in Table 2. At Step 2 of the analysis we observed facilitative effects of word frequency and concreteness, as in Experiment 1. We also found a significant inhibitory effect of item length. The main effects of Danger and Usefulness were not significant.

Table 2. Summary of Hierarchical Regression Analysis for Variables Predicting Visual Lexical Decision Time

<table>
<thead>
<tr>
<th>Variable(s)</th>
<th>df</th>
<th>$B (SE B)$</th>
<th>$\beta$</th>
<th>$F$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Between Subjects</td>
<td>30</td>
<td>---</td>
<td>---</td>
<td>18.57***</td>
</tr>
<tr>
<td>2. Control Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Concreteness</td>
<td>1</td>
<td>-0.2 (0.02)</td>
<td>-.17</td>
<td>103.29***</td>
</tr>
<tr>
<td>Animate/inanimate</td>
<td>1</td>
<td>-13.6 (11.49)</td>
<td>-.02</td>
<td>1.29</td>
</tr>
<tr>
<td>Neighborhood density</td>
<td>1</td>
<td>19.1 (17.88)</td>
<td>.03</td>
<td>1.14</td>
</tr>
<tr>
<td>Item length</td>
<td>1</td>
<td>35.1 (3.89)</td>
<td>.24</td>
<td>81.25***</td>
</tr>
<tr>
<td>Bigram frequency</td>
<td>1</td>
<td>-4.1 (4.14)</td>
<td>-.02</td>
<td>&lt; 1.0</td>
</tr>
<tr>
<td>Stress</td>
<td>1</td>
<td>-6.5 (14.63)</td>
<td>-.01</td>
<td>&lt; 1.0</td>
</tr>
<tr>
<td>Word frequency</td>
<td>1</td>
<td>-27.2 (3.68)</td>
<td>-.13</td>
<td>54.45***</td>
</tr>
<tr>
<td>3. Semantic dimensions</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Danger</td>
<td>1</td>
<td>-2.2 (2.09)</td>
<td>-.02</td>
<td>1.11</td>
</tr>
<tr>
<td>Usefulness</td>
<td>1</td>
<td>-1.2 (2.71)</td>
<td>-.01</td>
<td>&lt; 1.0</td>
</tr>
<tr>
<td>4. Danger x Usefulness</td>
<td>1</td>
<td>5.5 (1.13)</td>
<td>.17</td>
<td>23.63***</td>
</tr>
</tbody>
</table>

*** $p < .001$.

Crucially, though, the Danger x Usefulness interaction was again significant. Inspection of Figure 2 reveals that the nature of the interaction is the same as it was in Experiment 1 (although the main effect of Usefulness is gone). For words rated relatively low on Usefulness, higher levels of Danger are associated with faster RTs. For words rated relatively high on Usefulness, though, higher levels of Danger are associated with slower RTs. These slope differences are even more dramatic than they were with auditory presentation.
Figure 2. The slope of the relationship between lexical decision time and Danger, as a function of Usefulness. "Low" indicates a value one standard deviation below the mean and "High" indicates a value one standard deviation above the mean. The dimensions are continuous; Low and High values were used only for graphical purposes.

6 General discussion

A growing literature suggests that some aspects of semantic information are available before the unique identification of one word. The current study is an addition to this literature. It provided a clear replication of Wurm and Vakoch’s (2000) auditory lexical decision study, and extended that earlier work by showing that a very similar pattern of results emerges when the same stimuli are presented visually. Furthermore, the current study extends some of these findings in another way, in that effects such as concreteness were controlled for prior to the assessment of the semantic variables being examined.

Experiment 2 provided strong evidence that the Danger x Usefulness interaction is not tied specifically to auditory word recognition. The results are thus suggestive of a general word-recognition process, and in fact may
hint at an even more general information-processing mechanism. To explore this, future research will need to look beyond word recognition altogether and into other kinds of perceptual tasks. However, we consider Experiment 2 quite exploratory, primarily because the pseudowords were created (and some of the regressors were calculated) from an auditory framework. This perhaps explains some of the discrepancies across experiments. For example, the neighborhood density coefficient was very large and significant in Experiment 1, but small and not significant in Experiment 2. This makes sense, given that the density measure we used was based on auditory competitors, which introduces some degree of noise when compared to what the "proper" calculation would be. In addition, it could be that the Danger main effect would be significant in the visual experiment if more subjects were run; the regression coefficients (both unstandardized and standardized) are very similar in absolute size to those found in Experiment 1. Nevertheless, a replication of Experiment 2 will need to be done with visually-appropriate pseudowords, and some recalculated regressor values.

Several interpretational possibilities exist for the data pattern observed. One is that all information coming into the system is sent through a kind of perceptual filter, and classified (at least roughly) on Danger and Usefulness. The idea is that making these kinds of distinctions quickly has been a valuable thing to do throughout evolutionary history, and as a result, every kind of perceptual act passes through this sort of first-stage classification. The results of this classification would then be what we are observing in studies like the current one.

The embodied cognition approach is perhaps related to this, and represents another interesting possibility. The major idea behind embodied cognition is one of affordances: What kinds of sensory and motor acts can various objects be used for, given the bodies that we have and the world that we inhabit? Perception and other cognitive processes are constrained by the physical bodies that we have, and by what we can do with them. This leads to the prediction that we ought to see tight linkages between perception and action -- much tighter than are proposed in traditional models. Perhaps we are seeing evidence of that in the current study.

In one possible framework for interpreting the observed effects, we have proposed a two-part semantic analysis, with the first part being a very fast preliminary read based on Danger and Usefulness (or other dimensions that code similarly valuable information -- see Wurm & Vakoch, 1996; Wurm et al., in press; Vakoch & Wurm, 1997). Results of this first pass are integrated with the results of a second, more detailed analysis.

Research from many different areas is leading to a similar conclusion about semantic information being available (and having effects) at very early stages
of word recognition. The way people process morphologically complex words is affected by the transparency of the semantic relationship between the word constituents (e.g., Libben, 1998; Libben, Derwing, & de Almeida, 1999; Marslen-Wilson, Tyler, Waksler, & Oldser, 1994; Schreuder & Baayen, 1995; Wurm, 1997, 2000; Wurm & Ross, 2001), even though this transparency cannot be known for certain until the recognition process is completed. Van Petten, Coulson, Rubin, Plante, and Parks (1999) reported that N400 event-related potentials (ERPs) were different for stimulus words that were contextually appropriate and contextually inappropriate. These differences were evident 200 msec before the point at which the stimulus words could be uniquely identified. These authors concluded that "...semantic integration can begin to operate with only partial, incomplete information about word identity" (p. 394).

There is also recent work, both behavioral and psychophysiological, to suggest that affect, object recognition, and motor responses all interact. Chen and Bargh (1999) had participants either pull a lever toward themselves or push a lever away as a means of classifying visually-presented words as either good or bad. Participants responded to good words faster by pulling the lever than by pushing it, and they responded to bad words faster by pushing the lever away than by pulling it. A follow-up experiment was run in which the participants did not make evaluative classifications at all; their task was simply to pull the lever (or push it away, in another condition) whenever a word was displayed. Here, too, pull responses were faster for good words and push responses were faster for bad words. More recently, Schupp, Junghöfer, Weiße, and Hamm (2003) concluded on the basis of ERP patterns for participants looking at positive, neutral, and negative images, that "...the affect system not only modulates motor output (i.e., favoring approach or avoidance dispositions), but already operates at an early level of sensory encoding" (p. 7).

Many kinds of models could in principle be modified to accommodate the current findings. Weights on semantic dimensions could be used to change the resting activation levels of words or to modify the threshold values for word recognition. It would also be possible to order words according to their dimension weights, in the same way that some models have bins of words ordered by their frequencies of occurrence. Semantic information could also be incorporated as a higher-level information source that feeds down to earlier processing levels in a network.

It might also be possible to model early semantic effects without this semantic information being stored as part of a word's lexical entry. A model of semantic categorization proposed by Forster and Hector (2002) provides an interesting starting point. They asked participants to quickly decide whether something was an animal or not. Participants performed this task more poorly when
pseudowords like *turtle* had animal neighbors. Forster and Hector (2002) proposed a cascaded system that continuously monitors activation in semantic features (e.g., "animalness," or Danger).

This is an attractive possibility in that semantic variables can clearly have effects prior to the unique identification of a word. However, Wurm et al. (2003) and unpublished data from our lab suggest that these same effects are present in naming, and it is not clear that Forster and Hector's (2002) account would predict this. Unless required for the task at hand (e.g., semantic categorization), it is not obvious why the perceptual system would monitor activation in semantic features. The cascaded architecture is interesting, though, and should be explored further.

Gaskell and Marslen-Wilson's (1997) model, in which semantics and phonology are activated in parallel as part of a single process, is perhaps even more promising. It would need to be modified to specifically include the Danger and Usefulness effects, but such a modification seems relatively straightforward.

The methodology used in the current study is easy to use. We find that participants will willingly rate any words on any semantic dimensions, including extremely abstract words, unfamiliar words, and even function words (Vakoch & Wurm, 1997). Therefore, this approach can be used to explore any kind of meaning spaces that someone might imagine. We favor a framework built around Danger and Usefulness because of the behavioral significance they seem to have.

A crucial part of our framework is currently underspecified: What, exactly, do we mean when by "semantic?" Most linguists would refer to effects like Danger and Usefulness not as semantic, but as pragmatic. However, these variables seem to be as semantic as most of what passes for semantics in the field of psycholinguistics; and a strong version of the claim we are making is that these effects are actually a part of the lexical entry of a given word -- that is, they are a part of what each word means. In addition, other approaches to word recognition have either ignored semantics altogether, or suffer from the same underspecification problem. A possible exception to this is represented by some co-occurrence approaches to meaning (Buchanan, Westbury, & Burgess, 2001; Burgess & Lund, 1997; McDonald & Shillcock, 2001). In these approaches a word's semantics is fully specified by the company it keeps, defined in one of several similar ways. These approaches seem extremely promising in that measures derived in this way are related to certain performance outcomes (e.g., visual lexical decision times). However, others in the field are debating the validity of this approach, finding it intuitively unsatisfying and incompatible with their assumptions about the nature of cognitive representation. Future research will have a great deal to say about
this interesting debate.

In addition to running a better-controlled version of Experiment 2, a number of different directions for this work seem worth pursuing. One idea is to gather physiological or imaging data from participants who are being exposed to stimuli that vary in Danger and Usefulness. This would allow us to see if these effects correlate with what is known about processing of emotion (e.g., fear) in the brain. Another idea is to explore these kinds of effects as a function of part-of-speech (verbs in particular might be interesting to examine). Finally, recent work suggests that classification of objects, even nonsense figures, as either good or bad takes place automatically and very quickly after stimulus onset (Duckworth, Bargh, Garcia, & Chaiken, 2002). We therefore would also like to extend this work to look at other kinds of stimuli, such as pictures, environmental sounds, and pseudowords.

References


