# Mapping spoken words to meaning

**James S. Magnuson**
*University of Connecticut, Department of Psychological Sciences*

**PREPRINT OF CHAPTER:**
Magnuson, J. S. (2017). Mapping spoken words to meaning. In G. Gaskell & J. Mirkovic, *Speech Perception and Spoken Word Recognition* (pp. 76-96). New York: Routledge.

## ABSTRACT

This chapter provides a brief review of the nature of conceptual representations and how such representations are activated over time as spoken words are heard, including "pairwise" and "dimensional" empirical approaches, computational modeling, and the ubiquity of context effects. I conclude that while context and expectations have large impact on relative salience or activation of semantic features, such influences can only rarely override bottom-up priority, and that theoretical progress will require computational modeling.

The ultimate goal of language is to communicate *meaning*. As Marslen-Wilson (1987) put it, "to understand spoken language is to relate sound to meaning." As the acoustic events that make up a spoken word occur, listeners must map phonological forms to intended words in memory. Most research on spoken word processing focuses on form recognition – determining what words are activated in memory as a spoken word is heard, whether and how much they compete, and what factors promote or inhibit recognition. But models that end with abstract, phonological word forms ignore the primary function of words: accessing meaning and generating compositional meaning (based upon word combinations and syntactic structure). For many years, many psycholinguists assumed spoken word processing proceeded in modular stages, with exhaustive activation of words matching a phonological form, followed shortly by access to meaning and integration with syntax (e.g., Marslen-Wilson, 1987; Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979). Over the course of this chapter, we shall see that the notion of modular stages must be discarded; how quickly semantic access can be detected depends on the strength of expectations (based on semantic features or context), although human spoken language processing appears to conform to the principles of *bottom-up priority* and *delayed commitment* (Luce & Cluff, 1998; Marr, 1982): top-down expectations are not sufficient to drive full access or recognition without bottom-up support. I will begin by reviewing the nature of conceptual knowledge, then turn to the semantic details associated with individual words and how they are accessed in word recognition, discuss the very few computational models of spoken word processing that incorporate semantics, and conclude with a review of how word-level semantics interact in sentence and discourse contexts. Space is short, and each section is necessarily brief and selective.

## Nature of conceptual representations

Before we can address what semantic representations are associated with words, we must consider theories of semantic representations. Important concerns regarding distinctions between semantic knowledge and memory (world knowledge) vs. lexical knowledge are beyond the scope of this chapter; we will dive straight into proposals for representations of lexical knowledge.

Classical accounts based on necessary and sufficient features were rejected in the latter half of the twentieth century since many if not most concepts cannot be defined in such a way (cf. Wittgenstein's [1953] example of GAME). Collins and Quillian (1969) proposed that semantic knowledge might be organized as taxonomic networks embodying positive aspects of classical accounts while providing a compact basis for generalizations. For example, a node for ANIMAL would branch into subordinate nodes such as BIRDs, MAMMALs, REPTILEs, etc. Any features associated with ANIMAL (HAS-SKIN, BREATHES) would be inherited by all subordinate nodes, just as subordinates of BIRD (RAPTOR, SONGBIRD) would inherit all properties associated with BIRD. However, core predictions of the taxonomic approach were quickly falsified (e.g., time needed to evaluate "a pig is a mammal" should be proportional to nodes that must be traversed, but participants confirm "a pig is an animal" [two nodes] faster than "a pig is a mammal" [one node]).

Collins and Loftus (1975) abandoned hierarchy for a network model where associated concepts were linked with strengths proportional to *association*. Association strengths can be quantified by asking many subjects to list words brought to mind by a target word (Nelson, McEvoy, & Schreiber, 2004). Association norms provide reliable predictions of priming and other behavior, and there is a rich tradition in psycholinguistics of exploiting associations to explore lexical access. For example, in *cross-modal semantic priming*, a series of spoken words are presented. Occasionally, a letter string is displayed, and participants make a lexical decision. Priming from auditory to printed words is used to infer what words the auditory word activated. While spoken words prime themselves and associates (i.e., a target word is processed faster when preceded by itself or an associate vs. an unrelated word), more subtle chains of priming can be observed. For example, hearing DOCK can prime CAT, indicating that as DOCK is heard, phonologically similar words are activated (e.g., DOG) and in turn spread activation to their associates (Marslen-Wilson, 1989). However, the bases of associations are unclear (McRae, de Sa, & Seidenberg, 1997), as they include items related by category/feature overlap (BANANA-APPLE), synonymy (COUCH-SOFA), antonymy (NIGHT-DAY), function (LOCK-KEY), co-occurrence (BREAD-BUTTER), events (BRIDE-WEDDING), idiomatic expressions (KICK-BUCKET), or combinations (CAT and DOG overlap in features, children [or adults under time pressure] are likely to call them opposites, one is a likely chaser of the other in a CHASE event, and occur in idiomatic expressions such as "fighting like/raining cats and dogs"). The Collins and Loftus approach was eventually rejected as too limited and circular; networks constructed from empirically observed association strengths can do little beyond redescribing those associations.

Rosch and colleagues proposed a radically different approach based on featural similarity (e.g., Rosch & Mervis, 1975). On their "prototype" view, the basis for a concept is a set of prototypical features abstracted from the members of a category – a frequency-weighted composite of experienced features (e.g., the BIRD prototype would be a combination of body size, body part sizes, color, singing/calling, and feeding preferences [etc.] that do not occur in any single subtype or instance of BIRD). This view motivated many experimental paradigms that greatly expanded the known phenomena related to concepts and categories. For example: when participants list features of a concept, they first list highly typical features, whether or not they are distinctive (e.g., "has fur" might be elicited for virtually any mammal); processing time in many tasks depends on how close the target concept is to its prototype; there is graded membership within a category and fuzzy boundaries between categories, such that WHALE is rated both a poor MAMMAL *and* a poor FISH; there is a "sweet spot" level of specification based on expertise – the "entry" or "basic level" – that most observers default to, e.g., for non-experts, GUITAR rather than INSTRUMENT or TELECASTER. However, prototype theory fails to account for within-category structure (covariation among properties) that human subjects are sensitive to (for example, that smaller-bodied birds are more likely to sing, and larger-bodied birds are more likely to eat fish). Such findings motivated *exemplar theory* (e.g., Medin & Schaeffer, 1978; Smith & Medin, 1981), which proposes that a system in which each exemplar of a category is stored in memory would account for everything prototype theory did, while preserving a basis for sensitivity to within-category structure.

Tremendous progress has been made in feature-based approaches thanks to work eliciting features for concepts from human subjects (e.g., McRae et al., 1997; Vigliocco, Vinson, Lewis, & Garrett, 2004). Elicited features can provide a basis for quantifying semantic distance between concepts in the form of *semantic feature vectors*. Such vectors have one element for every feature that was elicited for any concept. The representation for a concept is a vector with a 1 for any feature that was elicited for that concept, and a 0 for any feature that was never elicited for that concept. Semantic distance can then be operationalized as Euclidean or cosine distance between vectors. Such similarity spaces provide accurate predictions of semantic similarity, priming, and other aspects of behavior associated with word-level semantics, despite not explicitly encoding category structure or causal relations (some experiments are described in more detail below).

But some phenomena are difficult to accommodate within feature-based approaches, most notably, ad-hoc and goal-directed categories (Barsalou, 1983), such as "good vehicles for crossing the desert" or "good activities for kids when you've missed a connecting flight". Even when such goals are first encountered, people propose instances and features with the same robust and gradient structure around prototypical tendencies observed for established categories (Barsalou, 1985). A prototype or exemplar approach cannot readily accommodate these phenomena; both theories require experience with exemplars for a structured category to form. Motivated by such results, Murphy and Medin (1985) argued that similarity-based theories of categorization failed to provide an account of *conceptual coherence* — what makes "some groupings … informative, useful and efficient" – and also lacked a truly explanatory account of "intra- and inter-concept relations". They argued that concepts are based on *theories* — "mental 'explanations'" – of *why* a category exists (what links members, and their purposes and functions). For example, people know that an airplane made of whipped cream could not fly, while one made of plexiglass *might* be able to, and even young children agree that while a blender could be reconfigured into a toaster, a skunk cannot be reconfigured into a cat (Keil, 1989). But this "theory theory" replaces prototype abstraction or exemplar clustering with something mysterious; what would be the nature of "mental explanations" underlying categories on this view? What kind of general mechanism might generate them based on experience with exemplars? (For a review, see Murphy, 2002.)

Another possibility is that conceptual coherence could emerge from simple learning mechanisms that encode appropriate relations. McClelland and Rogers (2003) developed a simple feedforward network that maps *Items* (concepts) and a small set of core *Relations* (IS, CAN, HAS) to *Attributes* (features) via layers of hidden nodes. After training, the model exhibits human-like clustering in its learned representations as well as generalization (e.g., given a subset of features for a new concept, it "infers" [partially activates] features that have not yet been experienced based on "coherent covariation" of feature relations it has learned). In contrast to theory-theory, the model becomes sensitive to causal structure within categories without any understanding or theory-like propositional representations. While the model does not provide a full account of conceptual coherence, it is a promissory note for an approach that might be extended to do so.

Other crucial approaches include *embodied* or *grounded* theories of cognition, which stem from the *symbol grounding problem* (the idea that words cannot be "ungrounded" symbols that are defined by other words, but must to be linked to semantics outside a linguistic system; Harnad, 1990) and from empirical findings. On this view, representations are not amodal abstractions; instead modal (sensory, motoric) brain areas provide an important component of representations (Barsalou, 2008). The neural representation of REACH, for example, would include activation of areas involved in perceiving and/or performing reaching. Many studies have demonstrated that performance on linguistic tasks can be facilitated or inhibited depending on the compatibility of a linguistic stimulus with the motor action required to respond (Glenberg & Kaschak, 2003), suggesting that linguistic forms trigger modal activations that engage perception-action systems. Barsalou (1999) proposes that hearing or thinking about a concept induces a "perceptual simulation" via appropriate modal activations (see also Paivio, 1986). A growing body of neuroimaging studies are consistent with this view. In an elegant, comprehensive theoretical and empirical review, Meteyard, Cuadrado, Bahrami, and Vigliocco (2012) conclude that behavioral and neural data point towards Damasio's (1989) proposal for *convergence zones* — neural regions that respond to and activate a variety of modal and amodal semantic details. Delving more deeply into neural representations is beyond the scope of this chapter, but Gow (2012) provides a recent review that also introduces a novel theory of phonetic and semantic neural representation.

Finally, a crucial consideration is the distinction between concrete and abstract concepts. It may seem that neither feature-based nor grounded approaches would provide a basis for representing abstract concepts. While one can get surprisingly far with intuitive features for many abstract concepts (see, e.g., Plaut & Shallice, 1993), simple features for many concepts may be difficult to derive, and features for many abstract concepts may be context- or situation-specific, much like ad-hoc and goal-derived categories. While Barsalou and Weimar-Hastings (2005) have proposed that even abstract concepts could rely on situation-specific perceptual simulations, Crutch and Warrington (2005) proposed that concrete and abstract concepts depend upon "qualitatively different representations". Specifically, they proposed that concrete concepts are represented via semantic features, while meanings of abstract concepts are based on associations with other concepts. Andrews, Vigliocco and Vinson (2009) also propose a qualitative distinction, arguing that concrete concepts are based primarily on experiential learning (based on co-occurrence of language and physical objects and events) while abstract concepts are learned distributionally (i.e., based on co-occurrence statistics derivable from linguistic sources [speech or print], as in the Hyperspace Analog to Language [HAL, Lund & Burgess, 1996] or Latent Semantic Analysis [LSA, Landauer & Dumais, 1997]), although they propose that the two sources are intercorrelated (their Bayseian model exhibits benefits from having both sources that are not a simple effect of having more data, but follow from having two qualitatively distinct types of data). We will review results that bear on this issue below. With these preliminaries in hand, let us turn to empirical results in the processing of spoken words.

## Aspects of meaning activated by single spoken words

Experiments on semantic access from single spoken words[1] can be divided into two primary types that sometimes appear to provide conflicting data. *Pairwise approaches* examine effects for specific word pairs – e.g., does DOG prime CAT, or in a paradigm where multiple images are presented, do participants look to CAT when they are told to select DOG? *Lexical dimension approaches* examine processing for *sets* of words as a function of lexical variables, such as frequency of occurrence, phonological neighborhood size, or semantic variables such as how imageable a word is, how many semantic features it has, etc. We shall see that while convergent results are often found using these approaches, they can reveal different aspects of lexical access.[2]

**Pairwise approaches.** A foundational *pairwise approach* was reported by Moss, Ostrin, Tyler and Marslen-Wilson (1995), who examined multiple types of semantic relation effects in spoken word recognition. They used "category coordinate" pairs belonging to the same category that were also associated or nonassociated, and were natural or artifacts (e.g., associated natural category coordinates: DOG-CAT; non-associated artifact category coordinates: SPADE-RAKE). "Functionally related" pairs were also associated or not, and were related by "script" or instrumental relations (associated, instrument: HAMMER-NAIL; non-associated, script: RESTAURANT-WINE). There were also many pairs of unrelated filler words. Word-word or word-nonword pairs were presented and participants made lexical decisions on the second item. For spoken presentation, reliable effects of relatedness were found for all item classes, along with significant boosts for association and an interaction of relatedness and association (greater priming for associated than nonassociated pairs). Concerns about strategic effects that could arise from word pairs motivated a second experiment with lexical decisions for each item. The results replicated, though priming for nonassociates was greatly reduced. A third experiment used visual

---

[1] Note that the literature on semantic dimensions is vastly larger for written than spoken words. However, we cannot be certain that effects from studies with visual words will generalize to spoken words, given differences in how words are experienced in the two modalities (all at once for visual, but serially over time for spoken).

[2] A crucial preliminary issue is that whether one is able to detect semantic effects in an experiment depends on the type of task used. For example, subtle semantic effects are difficult to find using a lexical decision task (press "YES" if what you hear is a word, press "NO" if it is not a word). Among the problems with this task is that it is possible for responses to be generated based on familiarity rather than recognition, and some have argued that it may reflect post-perceptual processing (e.g., Balota & Chumbley, 1984). Tasks that explicitly tap semantic dimensions are, unsurprisingly, much more sensitive to semantic dimensions (e.g., artifact judgments [press "ARTIFICIAL" or "NATURAL"] or imageability/concreteness judgments [can you visualize this object, or can you touch this object]).

word presentation, and the results were markedly different. Functionally related pairs primed whether they were associated or not; category coordinate pairs only primed if they were associated; and script-related pairs did not prime whether they were associated or not. One possibility is that only the third study measured automatic priming, and that surviving effects indicate automatically activated dimensions. However, as Moss et al. argued, differences may have followed from variation in prime-target intervals (200 ms for the first experiment, 1000 ms for the second, variable for the third). Another possibility is that differences in the way spoken words and written words are experienced (segment-by-segment vs. all at once) could have influenced the outcomes. What is clear is that not all semantic relationships are equivalent, and some may be activated more strongly and/or quickly, with complex differences in visual and spoken modalities.

The advent of the *visual world paradigm* (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), or VWP, opened up new ways of looking at pairwise relations as well as lexical dimensions. Allopenna, Magnuson, and Tanenhaus (1998) used it to examine phonological competition (see also Spivey-Knowlton, 1996) and introduced the use of fixation proportions over time to provide estimates of word activations. Yee and Sedivy (2001, 2006) conducted the first pairwise explorations of semantic features using the VWP. In one study, they displayed pictures of semantically related items (primarily category coordinates, such as CAR-BIKE or PIANO-TRUMPET, but also some functionally related items like LOCK-KEY or HAMMER-NAIL) along with unrelated distractors. Given an instruction to touch a target item, participants were significantly more likely to look at functionally related than unrelated items. Follow-up analyses confirmed that results were robust when restricted to item sets where visual (shape) similarity of targets and semantic relatives was matched to target-unrelated similarity, or to non-associated target-relative pairs. A second experiment sought to eliminate any possibility of the results being driven by strategy or visual properties of displayed pictures by replacing targets with phonological cohorts (e.g., CARDS replaced CAR, LOGS replaced LOCK). The logic was that Allopenna et al. and Spivey-Knowlton had demonstrated robust co-activation of onset competitors, and robustly activated items should spread activation to semantic relatives. Indeed, LOGS activated KEY (for example) significantly, though not as much as LOCK had.

In a similar study, Huettig and Altmann (2005) examined fixations over time as participants heard sentences mentioning a critical item as they viewed a display with pictures of the target (e.g., PIANO) and either (a) three unrelated distrators, (b) a category coordinate of the target (e.g., TRUMPET) and three unrelated distractors, or (c) both the target and the category competitor and two unrelated distractors. Participants were more likely to fixate targets and competitors than unrelated items, with earlier and higher peak fixation proportions for targets than competitors. Huettig and Altmann argued that fixations in the VWP reflect a variety of bottom-up (spoken word forms and pictures) and top-down activations (e.g., activation of TRUMPET from a MUSICAL INSTRUMENT feature activated by PIANO, or spreading activation among category members).

It seems implausible that visual displays would not impact processing in the VWP, despite the shape-similarity control analysis of Yee and Sedivy (2006). Indeed, Dahan and Tanenhaus (2005) found that items with shape similarity but no semantic relation to a target word (e.g., SNAKE-ROPE, CATERPILLAR-TRAIN) were fixated reliably more than unrelated items. Dahan and Tanenhaus argued that the VWP does not involve implicit naming of displayed objects; their interpretation was that if participants activated phonological forms of displayed pictures prior to the spoken target, one would not expect fixations to visual (shape) competitors. Fixations to those competitors suggest that visual features were activated by hearing the target word.

Huettig and Altmann (2007) came to a very different conclusion. Their key experiment involved shape-similar pairs like SNAKE-CABLE, with two display types (three unrelated pictures displayed either with a target [SNAKE] or a shape competitor of the target [CABLE]), and two sentence types: neutral (e.g., 'In the beginning, the man watched closely, but then he looked at the snake and realized that it was harmless'), or biasing (e.g., with 'man' replaced with 'zookeeper'). Competitor displays were paired only with biasing sentences. Participants were more likely to fixate competitors than unrelated items. They found a very small but reliable "anticipation" effect in the biasing condition: subjects were reliably more likely to be fixating targets by word onset. Contra Dahan and Tanenhaus, they concluded that fixations in the VWP are not a pure reflection of spoken words activating visual features which then guide saccades; rather, they argued that item names are activated when items are viewed, and spoken words reactivate or boost those picture-based activations – otherwise, there should have also been anticipatory looks to the shape-similar competitor.

However, there are several reasons that further experimentation would be needed to confirm this. First, they used exceptionally long preview times; pictures preceded sentences by 1 second, and target word onsets were approximately 5 seconds after picture presentation. Long previews allow substantial time for strategic exploration or subvocal naming of pictures. Second, key biasing context words (e.g., zookeeper) would not activate targets specifically, but rather a target's category (e.g., although a zoo animal would be expected, the cloze probability for SNAKE specifically would be extremely low). One could not expect the category to activate perceptual features of all category members (e.g., ANIMAL would not be expected to activate shape features for SNAKE, such as LONG and CYLINDRICAL). Finally, substantial time between key context words and target words (~2 seconds) would allow fixations to each object, allowing one to guess that, e.g., the one animal would likely be the target. Very high fixation probabilities for shape competitors *despite* participants having so much time to examine the display reinforces the Dahan and Tanenhaus conclusion that hearing a word strongly activates perceptual features of its referent; if subjects *knew* what the pictures were and had activated their names, why else should hearing SNAKE activate its visual features sufficiently to drive fixations to the shape competitor?

Yee and colleagues have conducted several additional studies on semantic activation using the VWP. Yee, Overton, and Thompson-Schill (2009) compared activations for associated word pairs that were weakly vs. strongly semantically related. They found robust competition for strongly semantically similar pairs, but not for weakly similar pairs, suggesting that in the VWP, words activate distributed semantic features rather than simply spreading activation among associated words.

Yee, Huffstetler, and Thompson-Schill (2011) compared competition between pairs related by shape (FRISBEE-PIZZA) or function (TAPE-GLUE). An elegant aspect of their design was shape competitors were similar to targets in their *prototypical* shape (round, for FRISBEE-PIZZA), but shape competitors were displayed with non-canonical shape (e.g., a pizza *slice* was displayed). With 1000 ms preview, robust, sustained fixations to shape competitors emerged early, while fixations to function competitors were weak and late (but still significantly greater than for unrelated items). With 2000 ms preview, the pattern flipped. Yee et al. suggested that increasing preview duration allowed them to tap the system later, and therefore, shape features must be activated early and function features must be activated late. This only makes sense if we assume that each displayed item is processed in some way during the preview, and not that the longer preview affords time for strategies to emerge – e.g., for subjects to notice relations among items. However, even if there was a strategic element in play, the complete flip of shape vs. function dominance with preview time is consistent with the idea that function features are less readily activated.

Yee, Chrysikou, Hoffman, and Thompson-Schill (2013) asked participants to make concreteness judgments for spoken words while performing either a concurrent manual task or a mental rotation task with a foot-based response. Manual tasks interfered specifically with items participants have more experience touching (e.g., pencils vs. tigers), while the mental rotation task interfered with both. The growing body of research on grounded/embodied cognition demonstrating sensory-motor interactions with language processing challenges any notion of abstract, amodal lexical symbols in the brain. While such interactions are often taken as *prima facie* evidence that neural bases for action and perception are integral to concept representations, caution is warranted. One might argue that all we can be confident of is that interaction is evidence of interaction – alternative explanations, such as "simple" spread of activation must be ruled out. On the other hand, two other theoretical principles are relevant. First, sensory-motor interactions satisfy criteria for Garner (1974) interference, where irrelevant variation in one dimension (e.g., type of motor response) disrupts processing of another (e.g., words), which is a longstanding test for integrality of representational dimensions. Second, on radical distributed perspectives (e.g., Elman, 2009; Spivey, 2007), there is no mental lexicon or discrete representations of words; instead, the representation of a word is distributed neural activity over the entire brain as the word is processed (and so changes dynamically over time). That said, as Hickock (2014) reminds us, "neurons compute." Once a concept is represented neurally, to a first approximation, it does not matter where those neurons are or what other concepts, percepts, or actions they may also help represent – a neural code in the absence of its referent is still a representation, no matter how closely it resembles codes active for perception or action. On the other hand, the specific neurons and circuits involved in conceptual memory and their proximity to or identity with those active in perception and action may afford important functional advantages in terms of speed of processing and preparation for action.

Finally, let's reconsider the possible differences in representations of abstract and concrete concepts. For obvious reasons, VWP studies focus on concrete objects. But there have been at least two attempts to use the VWP to address whether abstract and concrete objects depend on fundamentally different mental representations. As mentioned in the previous section, Crutch and Warrington (2005) proposed the *Qualitatively Different Representations* (QDR) hypothesis, motivated by patient studies finding interference for associated abstract words (WAR-PEACE), but not for associated concrete words (NAIL-HAMMER), and interference for semantically related concrete words (NAIL-SCREW) but not semantically related abstract words (WAR-CONFLICT). Crutch and Warrington argue that this dissociation suggests that concrete words are organized by category relations, while abstract words are organized by association (though it is not clear how associations provide a sufficient basis for understanding abstract concepts). Crutch and Warrington dismiss prior computational work arguing that differences between abstract and concrete words follow from relative sparsity of available features (Plaut & Shallice, 1993), since features for those simulations were intuited rather than empirically derived. Duñabeitia, Avilés, Afonso, Scheepers and Carreiras (2009) posited that it follows from QDR that hearing/seeing abstract words should activate primarily associated concepts, while hearing/seeing concrete words should primarily activate semantically similar concepts, and, crucially, abstract words should activate associates more rapidly than concrete words should, given the primacy of association for abstract words. In critical trials in a VWP study, Duñabeitia et al. asked participants to click on pictures that "best corresponded" to spoken words. "Best correspondents" were concrete associates of abstract (SMELL) or concrete (CRIB) spoken words (associates for these examples were NOSE and BABY). Indeed, associates of abstract words were fixated earlier and more than associates of concrete words. Duñabeitia et al. concluded that the QDR hypothesis was strongly supported.

Magnuson, Brozdowski and Gordils (in preparation; Brozdowski, Gordils & Magnuson 2013a) revisited this result, given concerns about Duñabeitia et al.'s metalinguistic task (clicking on "best correspondents", which could motivate various strategies) and materials (many "abstract" words had relatively low imageability, but were still concrete, such as PASSENGER). In order to use only concrete spoken targets in the VWP, we used the phonologically-mediated paradigm of Yee and Sedivy (2006). In critical displays, an instruction might be “click on the fist”. Distractors included one unrelated item, and two phonologically-linked associates: CLOTHING (related to FIST’s abstract phonological cohort, FIT) and OCEAN (related to FIST’s concrete phonological cohort, FISH). Contra Duñabeitia et al., associates of concrete phonological cohorts were fixated slightly but reliably more than those of abstract cohorts or unrelated items (which did not differ reliably from each other). In a follow-up experiment, pictures were replaced with printed words, allowing us to use both concrete and abstract items, and the primary results replicated. Either there is not a privileged route for abstract associations, or differences between highly abstract and highly concrete objects are quantitative rather than qualitative.

**Lexical dimensions approaches.** In a foundational study of *lexical dimensions* in the auditory domain, Tyler, Voice and Moss (2000) examined interactions of imageability and phonological density. In lexical decision and naming tasks, high-imageability words were processed faster than low-imageability words. In naming, there was an interaction with phonological cohort size (two words are cohorts [in the same recognition *cohort* according to the Cohort model; Marslen-Wilson, 1989] if they overlap in at least the first two segments): there was a reliable effect of imageability (faster responses for high imageability) for large but not small cohorts. Following arguments from Plaut and Shallice (1993), Tyler et al. suggested that imageability speeds processing because high imageability words tend to have more semantic features, which in Plaut and Shallice’s computational model promotes stronger basins of attraction. The interaction with cohort size is consistent with many prior results in which the effect of one variable is more pronounced at weaker levels of a second variable. For example, orthographic regularity interacts with word frequency, such that regularity effects are much more pronounced for low frequency words than high frequency words (e.g., Andrews, 1982), presumably because high frequency words approach a ceiling level of processing facility based on how often they are experienced. Because of activation of phonological competitors, words with large cohorts are expected to be more difficult to process than those with small cohorts, providing an opportunity for imageability effects to manifest. Tyler et al. argued that the interaction of imageability with cohort size indicated continuous, cascaded integration of semantic and phonological activations.

Another important consideration is what happens when a spoken or written word form is lexically ambiguous because it maps onto multiple unrelated meanings (homophony, as in BARK) or multiple related meanings (polysemy, as in TWIST). Rodd, Gaskell, and Marslen-Wilson (2002) pointed out that previous claims of an ambiguity advantage had not controlled for ambiguity due to homophony vs. polysemy. In visual and auditory lexical decision studies, they found inhibitory effects of homophony, but facilitatory effects of polysemy. Mirman, Strauss, Dixon and Magnuson (2010) followed up this work with lexical decision and VWP studies comparing nouns with noun homophones (e.g., DECK [cards], DECK [boat]), verb homophones (e.g., SWALLOW [bird], to SWALLOW), and unambiguous nouns equated on numerous lexical dimensions (frequency, neighborhood density, cohort density, uniqueness point, number of syllables, length, and recorded duration). In lexical decision, responses to nouns with noun homophones were reliably slower than response to nouns with verb homophones, while responses to unambiguous nouns were reliably faster than responses to both types of homophone. In the VWP, targets were presented among three unrelated distractors. Response time patterns converged with lexical decision results, and participants fixated pictures of unambiguous nouns more quickly than noun-verb homophones, which they fixated more quickly than noun-noun homophones (mean fixation proportions showed the same pattern). In the context of Rodd et al.’s (2002) finding of facilitation for polysemy, we interpreted our results as indicating a U-shaped function: high similarity with overlapping representations (polysemy) facilitates processing, but when overlap is low or absent (as for homophones with distinct meanings), lower representational distance (our noun-noun case) leads to greater competition, with decreasing competition for less overlap (noun-verb homophones) and even less for unambiguous items. We shall see below that this supposition may be difficult to reconcile with other effects of semantic distance.

Sajin and Connine (2014) extended findings from visual word recognition (Pexman, Holyk, & Monfils, 2003) that semantic richness, as indexed by *number of features* (NOF), speeds processing. NOF was operationalized using feature norms collected by McRae and colleagues (McRae, Cree, Seidenberg, & McNorgan, 2005; McRae et al., 1997). In lexical decision and VWP studies, words with relatively high NOF were processed more quickly than words with lower NOF for clear speech, and words with high NOF were less affected by the addition of background "babble".

Mirman and Magnuson (2008) used the same semantic features to investigate semantic *neighborhood*, in analogy to phonological neighborhood (using visual presentation and lexical decision or animacy judgment). As mentioned earlier, using the McRae et al. norms, concepts can be represented as vectors of 1s and 0s indicating which features participants listed for each concept. Distance between vectors can be captured with vector cosine. Having relatively many (2 or more) near neighbors (cosine > .5) predicted relatively slower processing, while having relatively many (200 or more) distant neighbors (0 < cosine < .25) predicted faster processing. We interpreted this in terms of attractor dynamics; having many distant neighbors can be conceptualized as a creating a region of many attractors that collectively create a large basin of attraction that speeds initial trajectories towards a general "region" of state space, while having many near neighbors slows trajectories due to multiple strong nearby attractors creating strong competition within a (densely packed) region.

**Reconciling pairwise and lexical dimensions approaches**. Mirman and Magnuson (2009) followed up the Mirman and Magnuson (2008) study with an auditory VWP study employing a pairwise approach. On critical trials, a target (e.g., FALCON) and near semantic neighbor (e.g., PARTRIDGE) or distant semantic neighbor (e.g., OSTRICH) were displayed with two unrelated items. Fixation proportions were reliably higher for near than distant neighbors, and for distant neighbors than unrelated items. Notably, competition for distant neighbors was not detected in previous studies using priming paradigms, pointing to (a) the sensitivity of the VWP and (b) the power of the McRae et al. (1997, 2005) feature-based approach to semantic similarity. While apparent *competition* with distant neighbors in a pairwise study like this one may seem incompatible with facilitation in a lexical dimensions approach (i.e., Mirman & Magnuson, 2008), Mirman and Magnuson used the same attractor network (Cree, McRae, & McNorgan, 1999) to simulate both sets of results. Mirman and Magnuson (2008) found that the model predicted increases in indices of competition as near neighbors increased, and decreases as number of distant neighbors increased. Mirman and Magnuson (2009) found that the same model predicted graded decreases in priming as neighbor distance increases. This raises a crucial question: how can we differentiate simple co-activation from competition in the VWP?

A study combining pairwise and lexical dimension approaches is instructive on this point. Apfelbaum, Blumstein and McMurray (2011) used the VWP to examine activation of semantic associates (a mixture of different relationship types) as a function of frequency-weighted (phonological) neighborhood (FWN) size. Words were chosen from relatively high or low FWNs, and were displayed with a semantic relative and two unrelated items. They predicted that because words in low FWNs are subject to less phonological competition, they would activate more quickly and thus be more likely to robustly activate semantic associates than words in high FWNs (the same logic behind the Tyler et al., 2000, study reviewed above – the effect of one variable should be most pronounced at a "difficult" level of another variable). While they did find greater fixations for semantic relatives in the low-density condition, there was an unanalyzed but strong trend for faster and greater fixation proportions for *high-FWN* targets than low-FWN. This would seem to contradict the explanation that semantic relatives were more strongly activated in the low-FWN condition (because low-density targets should activate more quickly than high FWN targets). However, even classes of items that tend to facilitate one another *must* compete when they are simultaneously displayed in the VWP, minimally at the level of decision processes guiding gaze selection and mouse or button presses since only one item can be selected at any instant for gaze or other behavioral action. So if a low-FWN target spreads greater activation to a *displayed* semantic relative and fixation proportions are proportional to activation, that semantic relative will attract more fixations. Since the participant can fixate only one item at a time, this will necessarily reduce fixations to the low-FWN target. Isolating the FWN effect would require displaying targets varying in FWN among unrelated items, as Magnuson, Dixon, Aslin, and Tanenhaus (2007) did for word frequency and phonological density (and found robust effects of frequency, neighborhood density, and cohort density).

The same principles hold for *any* pairwise experimental paradigm; detecting co-activation of a pair of items that tend to facilitate one another implies competition at the decision level — as more items are activated, decisions should be slowed. Detecting facilitation requires that the overall facilitative influences outweigh competitive influences. Thus, facilitative pairwise relations (for items simultaneously displayed) must manifest as competition effects in the VWP, but appropriately designed lexical dimensions studies can detect both inhibition and facilitation. So while the VWP provides essential time course information, both pairwise and lexical dimension VWP designs may and must be employed to obtain full understanding of cooperative and competitive influences. Computational models are also essential, and are considered next.

## Computational models

The Cree et al. (1999) model used by Mirman and Magnuson (2008, 2009) is not a model of spoken word recognition; abstract form units are presented simultaneously and feed forward directly to feature units. Attractor dynamics follow from the settling process for recurrent connections among feature units. But the essential difference between visual and spoken word recognition is that spoken inputs are necessarily serial and have temporal extent. Very few models incorporate both the temporal extent of spoken words and representations of meaning; most SWR models are simply form recognition models (see Magnuson, Harris, & Mirman, 2012 and Mirman, this volume, for reviews), although there are three notable exceptions.

First, Gaskell and Marslen-Wilson (1997, 1999) introduced their *Distributed Cohort Model* (DCM), which is a simple recurrent network (SRN; Elman, 1990) mapping phonetic features (one time step per phoneme) to distributed phonological forms and to "semantics" — 50-element vectors with 50% of nodes set randomly to 1s and the rest to 0 (random sparse vectors are often used as convenience representations for semantics [Plaut, 1997], on the rationale that form-meaning mappings are nearly always arbitrary). Initial papers used lexicons of 36 to 276 words, and established that the DCM accounts for basic aspects of word recognition (Gaskell & Marslen-Wilson, 1997) and focused on the notion of *blending* in distributed models vs. inhibition-based competition in localist models. The idea is that when an input strongly supports two words (e.g., upon hearing /kæpt/, which could result in CAPTAIN or CAPTIVE), an output representation of distributed semantic features does not allow concept representations to compete as they do in localist representations with a separate node for each concept with inhibitory connections. Instead, the state of the network will blend the two semantic feature vectors (assuming for the sake of example that there are not other words strongly activated by the input /kæpt/).

Gaskell and Marslen-Wilson (2002) tested predictions with ambiguous word onsets predicted to result in semantic blending (e.g., /kæpt/) or unambiguous onsets (e.g., /garm/, consistent with one word, GARMENT). Ambiguous word fragments did not prime semantic relatives of consistent words (e.g., COMMANDER) but unambiguous fragments did. They provide a nuanced discussion of how such results follow naturally from distributed representations, but might not follow as easily from models using "direct competition" (dedicated inhibitory connections between localist nodes). Indeed, a localist model *could* simulate the same results. Without any modification, TRACE (McClelland & Elman, 1986) provides a basis for the prediction: an ambiguous fragment will activate multiple words, and competition among them will depress any consistent word's activation compared to the level of activation that would follow from a fragment consistent with only one word. We can predict from that difference that priming should be weaker in the former case. If we modified the model and added semantic representations (e.g., with phonemic nodes mapping to semantic features which in turn map to lemma nodes, cf. Magnuson et al., 2012), the priming effects could be modeled. Nonetheless, the DCM stands out as the most promising approach to both time course and meaning for modeling spoken words to date, although it has not been extended since 1999.

The second exception is an interactive activation and competition model developed by Chen and Mirman (2012). The model is conceived as having three layers: input forms (abstract letters or phonemes) connect reciprocally to word nodes which connect reciprocally to semantic features, although only form-word or word-semantic layers were used for any single simulation. As with the DCM, semantic features were abstract (as were their letter and phoneme representations). Simulations used very few input forms (0 or 7), word nodes (1-11), and semantic features (0, 10, 12, or 70), and simulations were designed to isolate essential principles of competition. Form-word simulations demonstrated how simultaneous presentation of input elements (as in print) vs. sequential input (as in speech) change competition dynamics (specifically, form neighbors facilitate in the former case but inhibit in the latter case; cf. Magnuson & Mirman, 2007). Word-semantic simulations demonstrated how distant semantic neighbors can speed processing while near neighbors inhibit processing (though it is not clear how the approach could account for the U-shaped facilitation/competition function proposed by Mirman et al. [2010] reviewed above). The simplicity of Chen and Mirman (2012) is a double-edged sword. It affords elegant demonstrations of disparate competition dynamics, but demonstration proofs of "micro-models" can only be considered promissory notes until they are shown to scale to, e.g., realistic phoneme and word inventories (Magnuson, in press).

The third exception is Plaut and Kello's (1999) ambitious attractor network model of the co-development of speech and word perception and production. The model learned to map over-time analogs of acoustic inputs to phonological forms and semantic features (again, arbitrary vectors), as well as to map from semantics back through phonology to an articulatory system. Unfortunately, this model has not been extended beyond some initial demonstration proofs, but is the most promising approach to date for advancing understanding not just of how sound is mapped to meaning, but how that mapping emerges over development.

Perhaps the two greatest challenges for models of spoken word recognition are greater input realism (ideally working with real speech rather than the abstractions all current models assume for convenience) and connection not just to word meaning, but to the syntactic and semantic complexities of sentence contexts. Next, let's consider how aspects of word meaning change dynamically in sentences and other contexts.

## Sentence and other context effects

Swinney (1979) and Tanenhaus et al. (1979) reported classic results that motivated a modular view of form and meaning processing that dominated psycholinguistics for nearly 15 years. In a neutral context, ambiguous homophones prime all consistent meanings (e.g., BUG primes both insect and spy). A context biased towards one meaning eventually primes the context-appropriate meaning much more strongly than other meanings, but both Swinney and Tanenhaus et al. found that such effects appeared to take a few hundred milliseconds to emerge. Such results suggested two stages of processing: exhaustive bottom-up form access followed by semantic-based selection. Shillcock and Bard (1993) demonstrated that immediate effects *can* be observed with very strong expectations established by syntactic and semantic constraints. For example, given a context such as *John said he didn't want to do the job, but his brother would,* no cross-modal priming was found for TIMBER (a semantic relative of WOOD) from WOULD even when the probe was displayed *during* the critical word. Using the VWP, Dahan and Tanenhaus (2004) found that sentence contexts predicting a specific lexical item seemed to eliminate phonological competition from context-inappropriate competitors. Similarly, Magnuson, Tanenhaus, and Aslin (2008) found that syntactic and pragmatic constraints (whether an adjective or noun was expected next given VWP display contents) immediately restricted competition – no phonological competition was observed from nouns when adjectives were expected, and vice-versa. Chambers, Tanenhaus, and Magnuson (2004) found that object affordances and implications of a verb and/or instrument constrained competition. For example, given an instruction to "pour the egg…" when there was a liquid egg and an egg still in its shell, fixation proportions to the egg-in-shell were no greater than to unrelated objects. Given a hook to manipulate objects in a workspace, and two whistles (only one of which was hookable, via a string attached to it), fixations to the unhookable whistle did not differ from fixation proportions to unrelated objects.

Such results demonstrate *preemption*: absence of competition expected from the bottom-up input, which is weak evidence for *anticipation*. Strong evidence for anticipation comes primarily from two sorts of studies: event-related potential experiments where large N400 responses indicate specific word expectations (e.g., at "an" in "The day was breezy so the boy went outside to fly an airplane", given the very strong expectation for the final noun phrase to be "a kite"; DeLong, Urbach, & Kutas, 2005), and VWP studies where expected items are fixated before they are named. For example, given a display with a boy, a piece of cake, and some toys, fixations were equivalent for inanimate objects when subjects heard "the boy will move the…" but fixations were directed towards the cake anticipatorily given "the boy will eat the…" (Altmann & Kamide, 1999). Kamide, Altmann, and Haywood (2003) reported more complex interactions of scenes and word meaning. Given two possible riders (GIRL, MAN) and two rideable objects (CAROUSEL, MOTORBIKE), fixation proportions favored expected relationships (greater fixations to CAROUSEL than MOTORBIKE given "the girl will ride the…", vice-versa for "the man will ride the…"), although expectations appear to be probabilistic (e.g., although most fixations were directed to CAROUSEL given "the girl will ride the…", fixations to MOTORBIKE were greater than to non-rideable items).

Ferretti, McRae, and Hatherall (2001) found additional support for strong impact of sentence context. After establishing that verbs prime typical agents, patients, instruments, and even specific features of patients (MANIPULATING primes NAÏVE), though not typical locations, they presented auditory sentence fragments such as "she arrested the…" (agent role filled, should only prime patient) or "she was arrested by the…" (patient role filled, should only prime agent) and then presented a visual target word appropriate for an agent or patient role (e.g., COP/CROOK) which participants had to name. Naming was facilitated only for expected roles. However, given that the VWP has proved more sensitive than a variety of cross-modal paradigms (e.g., Allopenna et al., 1998), Kukona, Fang, Aicher, Chen, and Magnuson (2011) explored similar sentence constraints using the VWP. In one experiment, every sentence was about something that "Toby" was doing (e.g., *Toby arrests the crook*). In critical displays, excellent agents and patients of the verb were displayed (e.g., CROOK and POLICEMAN). Despite the fact that the agent role was *always* filled by Toby (always pictured in the center of the display), equivalent "anticipatory" fixations were observed to both good patients and good agents (which would not be expected if participants make optimal use of context); fixations reliably favored patients only after the onset of the word naming the patient. A second experiment demonstrated reliable anticipatory preference with additional syntactic cues and time for constraints to have impact; all sentences were about things that happened to Toby (e.g., *Toby was arrested by the*…), and while initial fixations to the good patient and agent were equivalent as the verb was heard, a reliable anticipatory preference to fixate the agent emerged during the preposition. It seems that naming in the Ferretti et al. (2001) study measured the late dominance of the context-appropriate role, and was not sufficiently sensitive to pick up the weaker early co-activation of both roles.

Finally, meaning ascribed to objects in the world also includes something like discourse tags or situation models. Chambers and San Juan (2008) used the VWP and had participants follow a series of instructions with displayed items (e.g., *Move the chair to area 2; now move the chair to area 5; now return the chair to area 2*). An instruction beginning "now return" led to anticipatory eye movements to previously "visited" areas. Thus, recognition of spoken words entails accessing long-term knowledge of semantic features, but also situation-specific mappings between words, the environment, and discourse history.

## Conclusions

The implications for the mapping from spoken words to meaning is that exactly what dimensions of lexical representations are activated can vary tremendously with context. Normally, bottom-up details have substantial priority and lexical activation involves complex interactions between bottom-up and top-down constraints and context. Under extreme constraints (that is, unusually strong expectations), preemption, preactivation and anticipation can be observed for specific semantic classes (e.g., Altmann & Kamide, 1999) or lexical items (DeLong et al., 2005, Kukona et al., 2011), but such expectations do not override bottom-up input (which would lead to hallucination). Indeed, even under strong constraint, we activate semantic and syntactic classes that are incompatible with context given strong bottom-up cues (Kukona et al., 2011), which only unusually strong constraints appear to circumvent (Shillcock & Bard, 1993). Developing theories of the complex dynamics of how over-time speech input interacts with the phonological and semantic features of words and linguistic and nonlinguistic contexts – and how those interactions develop throughout the lifespan – will require the development of more comprehensive computational models, building on the foundational work of Chen and Mirman (2012), Cree et al. (1999), Elman (2009), Gaskell and Marslen-Wilson (1999), and Plaut and Kello (1999).

## References

Allopenna, P. D., Magnuson, J. S., and Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory and Language, 38, 419-439.

Altmann, G. T. M., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. Cognition, 73, 247–264.

Andrews, S. (1982). Phonological recoding: Is the regularity effect consistent? *Memory & Cognition, 10*, 565-575.

Andrews, M., Vigliocco, G., Vinson, D. P. (2009). Integrating experiential and distributional data to learn semantic representations. *Psychological Review*, 116(3), 463-498.

Apfelbaum, K., Blumstein, S.E., and McMurray, B. 2011. Semantic priming is affected by real-time phonological competition: Evidence for continuous cascading systems. *Psychonomic Bulletin & Review, 18*, 141-149.

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, 10, 340 – 357.

Barsalou, L.W. (1983). Ad hoc categories. Memory & Cognition, 11, 211-227.

Barsalou, L.W. (1985). Ideals, central tendency, and frequency of instantiation as determinants of graded structure in categories. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11, 629-654.

Barsalou, L.W. (1999). Perceptual symbol systems. Behavioral and Brain Sciences, 22, 577-609.

Barsalou, L.W. (2008). Grounded cognition. Annual Review of Psychology, 59, 617-645.

Barsalou, L.W., & Wiemer-Hastings, K. (2005). Situating abstract concepts. In D. Pecher and R. Zwaan (Eds.), Grounding cognition: The role of perception and action in memory, language, and thought (pp. 129-163). New York: Cambridge University Press.

Brozdowski, C. R., Gordils, J., & Magnuson, J. S. (2013, November). Contra the Qualitatively Different Representation Hypothesis (QDRH), concrete concepts activate associates faster than abstract concepts. Talk presented at the Psychonomic Society, Toronto.

Brozdowski, C. R., Gordils J., & Magnuson J. S. (in preparation). A concreteness advantage for concrete and abstract associations.

Chambers, C., & San Juan, V. (2008). Perception and presupposition in real-time language comprehension: Insights from anticipatory processing. Cognition, 108, 26–50.

Chambers, C. G., Tanenhaus, M. K., & Magnuson, J. S. (2004). Actions and affordances in syntactic ambiguity resolution. Journal of Experimental Psychology: Learning, Memory, & Cognition, 30, 687-696.

Chen, Q. & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review*, *119*(2), 417-430.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428.

Collins, A. M., & Quillian, M. R. (1969). Retrieval time from semantic memory. Journal of Learning and Verbal Behavior, 8, 240-247.

Cree, G. S., McRae, K., & McNorgan, C. (1999). An attractor model of lexical conceptual processing: Simulating semantic priming. Cognitive Science, 23(3), 371–414.

Crutch, S. J., Warrington, E. K. (2005). Abstract and concrete concepts have structurally different representational frameworks. *Brain* 128(3), 615-627

Dahan, D., & Tanenhaus, M. K. (2004). Continuous mapping from sound to meaning in spoken-language comprehension: Immediate effects of verb-based thematic constraints. Journal of Experimental Psychology: Learning, Memory, and Cognition, 30, 498-513.

Dahan, D., & Tanenhaus, M. K. (2005). Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. Psychonomic Bulletin & Review, 12, 453-459.

Damasio A. R. (1989). The brain binds entities and events by multiregional activation from convergence zones. *Neural Computation, 1*, 123-132.

Delong, K., Urbach, T., & Kutas, M. (2005). Probabilistic word pre- activation during language comprehension inferred from electrical brain activity. Nature Neuroscience, 8, 1117–1121.

Duñabeitia, J.A., Avilés, A., Afonso, O., Scheepers, C., & Carreiras, M. (2009). Qualitative differences in the representation of abstract versus concrete words: Evidence from the visual-world paradigm. *Cognition, 110*, 284-292.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179–211.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge without a lexicon. Cognitive Science, 33, 1–36.

Ferretti, T., McRae, K., & Hatherell, A. (2001). Integrating verbs, situation schemas, and thematic role concepts. *Journal of Memory and Language, 44*, 516–547.

Garner, W. R. (1974). *The processing of information and structure.* New York: Wiley.

Gaskell, M. G. & Marslen-Wilson, W.D. (1997). Integrating form and meaning: a distributed model of speech perception. *Language and Cognitive Processes,* 12, 613-656.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science, 23*, 439–462.

Gaskell, M. G. & Marslen-Wilson, W. D. (2002). Representation and competition in the perception of spoken words. *Cognitive Psychology,*45, 220-266.

Glenberg A. M. & Kaschak M. P. (2003). The body's contribution to language. In B. H. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 93-126). San Diego, CA: Academic Press.

Gow, D. W. (2012). The cortical organization of lexical knowledge: A dual lexicon model of spoken language processing. *Brain & Language, 121*, 273-288.

Harnad S. (1990). The symbol grounding problem. Physica D, 42: 335-346.

Hickock, G. (2014). *The Myth of Mirror Neurons: The Real Neuroscience of Communication and Cognition.* New York: Norton.

Huettig, F. & Altmann, G.T.M. (2005). Word meaning and the control of eye fixation: semantic competitor effects and the visual world paradigm. *Cognition*, 96(1), 23-32.

Huettig, F., and Altmann, G.T.M. (2007). Visual-shape competition and the control of eye fixation during the processing of unambiguous and ambiguous words. *Visual Cognition,* 15(8), 985-1018.

Kamide, Y., Altmann, G. T. M., & Haywood, S. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. Journal of Memory and Language, 49, 133–156.

Keil, F. (1989). Concepts, kinds, and cognitive development. Cambridge, MA: MIT Press.

Kukona, A., Fang, S., Aicher, K. A., Chen, H., & Magnuson, J. S. (2011). The time course of anticipatory constraint integration. *Cognition*, *119*, 23-42.

Landauer, T. K., & Dumais, S. T. (1997). A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review, 104*(2), 211–240.

Luce, P. A. & Cluff, M. S. (1998). Delayed commitment in spoken word recognition:

Evidence from cross-modal priming. *Perception & Psychophysics, 60*, 484-490.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. Ear and Hearing, 19, 1–36.

Lund, K., & Burgess, C. (1996). Producing high-dimensional semantic spaces from lexical co-occurrence. *Behavior Research Methods, Instru- ments, & Computers, 28*(2), 203–208.

Magnuson, J. S. (in press). Phoneme restoration and empirical coverage of interactive activation and adaptive resonance models of human speech processing. *Journal of the Acoustical Society of America*.

Magnuson, J. S., & Mirman, D. (2007). *Neighborhood effects in word recognition: It' s not where you live, it' s how you get home*. Paper presented at the 48th Annual Meeting of the Psychonomic Society, Long Beach, CA.

Magnuson, J. S., Dixon, J., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. Cognitive Science, 31, 133-156.

Magnuson, J. S., Mirman, D., & Harris, H. D. (2012). Computational models of spoken word recognition. In M. Spivey, K. McRae, & M. Joanisse (Eds.), *The Cambridge Handbook of Psycholinguistics* (pp. 76-103). Cambridge University Press.

Magnuson, J.S., Tanenhaus, M.K., & Aslin, R.N. (2008). Immediate effects of form-class constraints on spoken word recognition. Cognition, 108(3), 866-873.

Marr, D. (1982). *Vision*. San Francisco: W.H. Freeman.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. *Cognition*, 25, 71 – 102.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. Cognitive Psychology, 18, 1–86.

McClelland, J. L. & Rogers, T. T. (2003). The parallel distributed processing approach to semantic cognition. *Nature Reviews Neuroscience, 4,* 310-322.

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods, 37,* 547–559.

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General, 126*(2), 99–130.

Medin, D. L. & Schaeffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85*, 207-238.

Meteyard, L., Cuadrado, S. R., Bahrami, B., Vigliocco, G. (2012). Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex* 48(7), 788-804

Mirman, D. & Magnuson, J.S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. Journal of Experimental Psychology: Learning, Memory & Cognition, 34, 65-79.

Mirman, D. & Magnuson, J.S. (2009). Dynamics of activation of semantically similar concepts during spoken word recognition. *Memory & Cognition, 37*, 1026-1039.

Mirman, D., Strauss, T.J., Dixon, J.A., & Magnuson, J.S. (2010). Effect of representational distance between meanings on recognition of ambiguous spoken words. *Cognitive Science, 34*, 161-173.

Mirman, this volume

Moss, H.E., Ostrin, R.K., Tyler. L.K., & Marslen-Wilson, W.D. (1995). Accessing different types of lexical semantic information: Evidence from priming. Journal of Experimental Psychology: Learning, Memory & Cognition, 21(4), 863-883.

Murphy, G. L. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological Review, 92*, 289-316.

Nelson, D. L., McEvoy, C. L., & Schreiber, T. A. (2004). The University of South Florida free association, rhyme, and word fragment norms. *Behavior Research Methods, Instruments, & Computers, 36*(3), 402– 407.

Paivio A. (1986). *Mental representations: A dual-coding approach*. New York: Oxford University Press.

Pexman, P. M., Holyk, G. G., & Monfils, M.-H. (2003). Number of features effects and semantic processing. *Memory & Cognition, 31,* 842-855.

Plaut, D.C., and Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377-500.

Plaut, D. C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexical decision. *Language and Cognitive Processes, 12,* 765–805.

Plaut, D. C. & Kello, C. T. (1999). The emergence of phonology from the interplay of speech comprehension and production: A distributed connectionist approach. In B. MacWhinney (Ed.), *The Emergence of Language* (pp. 381-415). Mahwah, NJ: Erlbaum.

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2002). Making sense out of ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, *46*(2), 245-266.

Rosch, E. & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7,* 573-605.

Sajin, S. M. & Connine, C. M. (2014). Semantic richness: The role of semantic features in processing spoken words. *Journal of Memory & Language*, *70*, 13-35.

Shillcock, R., & Bard, E. G. (1993). Modularity and the processing of closed- class words. In G. T. M. Altmann & R. Shillcock (Eds.), Cognitive models of speech processing: The second Sperlonga meeting (pp. 163–185). Mahwah, NJ: Lawrence Erlbaum Associates.

Smith, E. E. & Medin, D. L. (1981). *Categories and Concepts*. Cambridge, MA: Harvard University Press.

Spivey, M. J. (2007). The continuity of mind. New York: Oxford University Press.

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

Tanenhaus, M. K., Leiman, J., & Seidenberg, M. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. Journal of Verbal Learning and Verbal Behavior, 18, 427–440.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. Science, 268, 1632–1634.

Tyler, L. K., Voice, J. K., & Moss, H. E. (2000). The interaction of meaning and sound in spoken word recognition. Psychonomic Bulletin & Review, 7, 320–326.

Vigliocco, G., Vinson, D. P., Lewis, W., Garrett, M. F. (2004). Representing the meanings of object and action words: The featural and unitary semantic space hypothesis. *Cognitive Psychology* 48(4), 422-488.

Wittgenstein, L. (1953). *Philosophical Investigations.* New York: MacMillan.

Yee, E., & Sedivy, J. (2001). Using eye movements to track the spread of semantic activation during spoken word recognition. Paper presented to the 13th annual CUNY sentence processing conference, Philadelphia.

Yee, E., & Sedivy, J. C. (2006). Eye movements to pictures reveal transient semantic activation during spoken word recognition. Journal of Experimental Psychology: Learning, Memory, and Cognition, 32, 1–14.

Yee, E., Chrysikou, E., Hoffman, E., & Thompson-Schill, S.L. (2013). Manual Experience Shapes Object Representation.*Psychological Science*, *24 (6)*, 909-919.

Yee, E., Huffstetler, S., & Thompson-Schill, S.L. (2011). Function follows form: Activation of shape and function features during object identification. *Journal of Experimental Psychology: General*, *140 (3)*, 348-363.

Yee, E., Overton, E., & Thompson-Schill, S.L. (2009). Looking for meaning: Eye movements are sensitive to overlapping semantic features, not association. *Psychonomic Bulletin and Review*, 16(5), 869-874.