Mechanisms of Semantic Ambiguity Resolution: Insights from Speech Perception

Daniel Mirman

Abstract   The speech signal is inherently ambiguous and all computational and behavioral research on speech perception has implicitly or explicitly investigated the mechanism of resolution of this ambiguity. It is clear that context and prior probability (i.e., frequency) play central roles in resolving ambiguities between possible speech sounds and spoken words (speech perception) as well as between meanings and senses of a word (semantic ambiguity resolution). However, the mechanisms of these effects are still under debate. Recent advances in understanding context and frequency effects in speech perception suggest promising approaches to investigating semantic ambiguity resolution. This review begins by motivating the use of insights from speech perception to understand the mechanisms of semantic ambiguity resolution. Key to this motivation is the description of the structural similarity between the two domains with a focus on two parallel sets of findings: context strength effects, and an attractor dynamics account for the contrasting patterns of inhibition and facilitation due to ambiguity. The main part of the review then discusses three recent, influential sets of findings in speech perception, which suggest that (1) top-down contextual and bottom-up perceptual information interact to mutually constrain processing of ambiguities, (2) word frequency influences on-line access, rather than response biases or resting levels, and (3) interactive integration of top-down and bottom-up information is optimal given the noisy, yet highly constrained nature of real-world communication, despite the possible consequence of illusory perceptions. These findings and the empirical methods behind them provide auspicious future directions for the study of semantic ambiguity resolution.

D. Mirman (✉)
Department of Psychology, University of Connecticut, Storrs, CT 06269, USA
e-mail: daniel.mirman@uconn.edu

1 Introduction

Successful communication requires resolution of ambiguities at every level of language processing, including phonological, lexical/semantic, syntactic, and text or discourse levels. Researchers have generally investigated ambiguity resolution at each level independently. In an effort to overcome this proliferation of different proposed mechanisms and representations, MacDonald et al. (1994) used general constraint satisfaction principles to integrate semantic and syntactic ambiguity resolution. Similarly, it may be possible to integrate semantic ambiguity resolution and phonological ambiguity resolution (i.e., speech perception) within a single set of mechanisms and thus make progress toward understanding each of the domains and language processing in general. Semantic ambiguity refers to the fact that many words have multiple senses, and often completely different meanings. One of the challenges of theories of word recognition is to explain how context, prior experience, and other factors are combined to resolve this ambiguity.

Phonetic ambiguity is an intrinsic property of the speech signal: due to articulatory dynamics and differences between speakers, invariant cues have not been found that identify the intended linguistic units from the sounds a listener hears (this is called the *lack of invariance problem*). Depending on articulatory context and the size, gender, dialect, accent, or even speech style of the speaker, a single acoustic stimulus can be identified as many different phonemes and a single phoneme can have many different physical realizations (a many-to-many mapping). Consequently the speech signal is inherently ambiguous and listeners must use contextual information to resolve this ambiguity; although the search for invariant cues (Fowler 2006) or compensatory perceptual mechanisms (Holt and Lotto 2002) continues to be an active line of research. The role of word-level knowledge in resolving this perceptual ambiguity has been the focus of a heated debate. Many empirical and computational investigations have examined whether a listener's knowledge of the words of his/her language has a direct influence on how (s)he hears individual speech sounds or whether this influence occurs at a post-perceptual decision level.

In both the semantic ambiguity resolution domain and the speech perception domain a central debate has concerned the mechanism by which bottom-up perceptual and top-down contextual information are integrated. Many of the issues are computationally isometric, thus theories regarding the cognitive architecture of semantic ambiguity resolution can be informed by theoretical developments in the domain of speech perception. That is, contextual influences on activation of and competition among multiple meanings (i.e., semantic ambiguity resolution) parallel lexical influences on activation of and competition among multiple phonemes (i.e., speech perception). Interactive processing models, such as the TRACE model of speech perception (McClelland and Elman 1986), provide the best account of these effects (McClelland et al. 2006). Interactive processing is based on the intuition that cognitive processing will be most effective if all processing levels mutually constrain each other and such
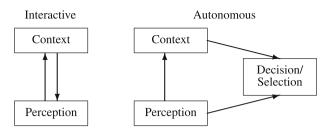
Fig. 1 A schematic of interactive (left) and autonomous (right) models. In interactive models, processing levels interact through bi-directional information flow. In autonomous models, contextual and perceptual information is integrated at a separate decision stage

models have been shown to implement optimal inference (Movellan and McClelland 2001). Interactivity is typically implemented by allowing both feedforward and feedback connections between different processing levels, although recurrent connections (Elman 1990) are another approach to interactive integration of context and perception (see Magnuson et al. 2008 for a review of computational models of spoken word recognition). Interactive models stand in contrast to autonomous models such as the Fuzzy Logical Model of Perception (Massaro 1998) and Merge (Norris et al. 2000). These models derive from the intuition that encapsulation of cognitive subsystems or processing levels is necessary for accurate and efficient processing (Fodor 1983). Autonomous models reject feedback, but typically allow integration of information from different levels to occur at a post-perceptual decision stage.

Interactive and autonomous speech processing models have conceptual parallels in the semantic ambiguity resolution domain for both the interactive approach (Kawamoto 1993) and the autonomous approach (Twilley and Dixon 2000). Figure 1 shows schematic representations of interactive and autonomous models. In the case of speech perception, the perception level corresponds to bottom-up acoustic/auditory perceptual processing and the context level corresponds to lexical processing and knowledge. In the autonomous model there is an additional decision/selection level at which information from perception and context levels is combined in the service of response selection and decision making; in the interactive model the decision/selection processes are thought to arise from the perception level, with context level information integrated via direct feedback. In the case of semantic ambiguity resolution, the perception level corresponds to bottom-up word form specification and the context level corresponds to representation of sentence, discourse, and/or pragmatic contextual information. As in the speech perception case, autonomous models posit that the integration of perceptual and contextual information takes place at a separate meaning selection level (Twilley and Dixon 2000), whereas interactive models posit that meaning selection takes place at the word (perception) level under direct feedback influence from context.

In addition to theoretical parsimony, there is a structural reason for symmetry between lexical effects on phoneme recognition and contextual effects on semantic ambiguity resolution: in both cases context is built up over time, with a single time slice providing very little information, but the input as a whole providing strong constraints. Spoken words are composed of phonemes, which are heard serially, and individually

provide very little lexical information, but together specify a unique lexical context; similarly, sentences are composed of words, which are heard or read serially, and individual words provide relatively little information about the sentence, but together specify a unique sentence context. This structural similarity and the general principle of theoretical parsimony suggest that the same computational architecture should be used for semantic ambiguity resolution and speech perception (though computational similarity does not entail similarity at the level of neural implementation, Marr 1982). To that end, predictions from interactive and autonomous models of speech perception can be taken as working predictions for interactive and autonomous views of semantic ambiguity resolution. This review begins by establishing the structural similarity between speech perception and semantic ambiguity resolution, then discusses three controversial and influential sets of speech perception findings and their implications for semantic ambiguity resolution.

2 Parallels Between Semantic Ambiguity Resolution and Speech Perception

The speech signal is inherently ambiguous and phoneme recognition shares structural similarity with semantic ambiguity resolution. As a result, there are many parallels between the kinds of effects that arise in the two domains. The following sections describe two cases in which there are clear parallels between findings in semantic ambiguity resolution and speech perception.

2.1 Context Strength Effects

A robust finding is that the effect of context on semantic ambiguity resolution depends on the strength of the context. For example, both meanings of an ambiguous word (e.g., *straw*) become active in weak contexts such as 'If Joe buys the straw', but the context-appropriate meaning tends to become more active in contexts that are more constraining, such as 'The farmer buys the straw' (Seidenberg et al. 1982). In the domain of speech perception, lexical context strength is reflected by the target phoneme's position in a word. If the target occurs early in the word, the lexical context tends to be weak because at this point the stimulus could turn out to be many different words. If the target occurs late in the word, the lexical context tends to be stronger because the word is more fully specified (a similar manipulation of word position in a sentence has also been used to study context strength effects on semantic ambiguity resolution: e.g., Duffy et al. 1988). The importance of lexical context strength is even clearer when one considers possible word completions rather than strict position in the word. For example, the point at which there is only one possible word completion is called the uniqueness point, which can occur earlier or later in a word depending on the composition of the lexicon. The uniqueness point is the point at which the context effect becomes specific to a particular word and lexical effects tend to be strongest after the uniqueness point (Frauenfelder et al. 1990), though listeners are imperfect and uncertain about completions of a spoken stimulus that is truncated at the uniqueness point (Grosjean 1980; Tyler and Wessels 1983). There is typically no uniqueness point in semantic ambiguity resolution because individual sentence context is not

sufficient to narrow the options to a single word meaning, though in real-world cases the combination of discourse, pragmatic, and syntactic factors may combine to specify a single possible meaning.

Lexical influences on phoneme processing are evident in faster recognition of phonemes in words than nonwords (Rubin et al. 1976). The effect of context strength on this word advantage has been examined by comparing the size of the word–nonword difference at different points within a word. As a word is heard, the set of possible completions (and thus the possible upcoming phonemes) becomes smaller and smaller, thus the strength of contextual support for a particular phoneme becomes stronger and stronger. This is analogous to progression through a sentence, such that the set of possible sentence completions becomes smaller and smaller, making context strength stronger and stronger. Evidence from behavioral studies shows that the word advantage effect increases through the word (Frauenfelder et al. 1990; Pitt and Samuel 1995). In particular, the word advantage is much stronger after the uniqueness point than before it, though there is some evidence for smaller benefits even before the word can be uniquely identified (Pitt and Samuel 1995). These results suggest that contextual influences are integrated gradually throughout processing, but early on, when the context itself is ambiguous, these effects are small and not easily detected.

Another common and robust lexical effect on phoneme recognition is the bias to hear an ambiguous phoneme so that it forms a word (Ganong 1980; Pitt and Samuel 1993) or a more frequent word (Fox 1984). For example, by blending two natural speech tokens or by using synthetic speech, the experimenter creates a continuum of sounds from /k/ to /g/. Then the tokens are appended to opposing lexical contexts such as /rʌ/ (i.e., rug-ruk) and /stʌ/ (i.e., stug-stuck). The finding in behavioral studies is that ambiguous tokens—those in the middle of the continuum—are more likely to be heard as /g/ when preceded by /rʌ/ (to form the word *rug* rather than the nonword *ruk*) and as /k/ when preceded by /stʌ/ (to form the word *stuck* rather than the nonword *stug*). Context strength can be manipulated by word length and phoneme position within the word. Again the results point to a graded effect of context strength: the lexical bias is stronger in longer words (Pitt and Samuel 2006) and stronger for word-final than word-initial phonemes (Pitt and Samuel 1993).

That the lexical bias occurs at all for word-initial phonemes (Ganong 1980; Pitt and Samuel 1993) is an interesting result because it requires context effects to act backwards in time. This means that phonological information and ambiguity must be maintained in a memory trace that can be updated by later-occurring inputs. This comprises a phoneme-level parallel to garden-path type sentences that require revision of early interpretation based on later-occurring information. These findings underscore the similarities between processing at the phoneme–word interface and processing at the word–sentence interface.

The attentional state of the perceiver is another important type of context that has been relatively unexplored. For example, if a listener hears several random sequences of words, (s)he may stop attending to syntactic structure and show reduced syntactic effects on ambiguity resolution on subsequent trials. In the domain of speech perception, recent behavioral studies have shown that attention does modulate the strength of lexical effects on phoneme recognition: when lexical information was generally useful for task performance, lexical effects were stronger (Mirman et al. 2008).

Neurophysiological studies of visual attention show weaker neural responses to dis-attended information (e.g., less activation in a motion-selective brain region when participants were attending to stationary stimuli and ignoring moving stimuli, O'Craven et al. 1997). If reduction of attention to lexical information causes a decrease in the activation of lexical representations (by analogy to visual attention), then it would consequently cause a reduction in the strength of lexical effects (Mirman et al. 2008). Attention is a critical aspect of cognitive processing and exploring the role of attention in ambiguity resolution may help to apply general principles of cognitive processing to understanding ambiguity resolution.

2.2 Attractor Dynamics Account for Contrasting Patterns of Inhibition and Facilitation due to Ambiguity

Some studies have reported faster word recognition for words that are semantically ambiguous (Hino and Lupker 1996), but other studies have reported slower word recognition for words that are semantically ambiguous (Rayner 1998). These paradoxical results can be reconciled by distinguishing two types of ambiguity: homophony (unrelated word meanings) and polysemy (related word senses) (Frazier and Rayner 1990). When these two types of ambiguity are distinguished in behavioral studies, there is a polysemy advantage and a homophony disadvantage (Rodd et al. 2002). This result is consistent with a dynamic attractor-based model of language processing in which polysemy broadens attractor basins, thus facilitating word recognition and homophony creates competitor attractors, which delay word recognition (Rodd et al. 2004).

The attractor-based view of semantic ambiguity resolution is also supported by recent studies of semantic neighborhood density effects demonstrating inhibition by near neighbors and facilitation by distant neighbors (Mirman and Magnuson 2008). Neighborhood size is a general concept referring to the number of words that are similar to the target word along some dimension. Specifically, a semantic neighborhood is comprised of words that are similar in meaning (Buchanan et al. 2001; Mirman and Magnuson 2008), a phonological neighborhood is comprised of words that are similar in sound (Luce and Pisoni 1998), and an orthographic neighborhood is comprised of words that are similar in spelling (Coltheart et al. 1977). In the semantic domain, near neighbors are very similar concepts such as *sheep* and *lamb*, and distant neighbors are partially similar concepts such as *sheep* and *squirrel*. Mirman and Magnuson (2008) found that semantic near neighbors act as competing attractors, thus delaying word recognition, and semantic distant neighbors broaden attractor basins, thus facilitating word recognition. Facilitative neighborhood effects are thought to reflect familiarity (i.e., more frequent exposure to words that mean, sound, or look similar to the target word) and inhibitory neighborhood effects are thought to reflect competition between similar words. This contrast is analogous to the ambiguity contrast: polysemy is associated with familiarity and therefore facilitation and homophony is associated with competition and therefore inhibition. However, these accounts provide no a priori method of predicting whether familiarity facilitation or competitive inhibition will emerge. In contrast, an attractor-based model of semantic processing predicts both the contrasting effects of homophony and polysemy (Rodd et al. 2004) and near and distant neighbors (Mirman and Magnuson 2008).

In speech perception, a similar contrast has been found between inhibitory lexical neighborhood effects and facilitative phonotactic probability effects (Luce and Large 2001; Vitevitch and Luce 1999, 1998). A lexical neighborhood is typically defined as the set of all words that differ from the target word by the substitution, addition, or deletion of a single phoneme (Luce and Pisoni 1998), for example the lexical neighborhood for *cat* includes *bat*, *cot*, *cab*, *cast*, etc. Phonotactic probability (Vitevitch and Luce 1999) is typically computed by combining the probability of all the two-phoneme sequences (biphones) in the word, for example for *cat* this would be the probability of /kæ/ and /æt/. To account for the opposing effects of lexical neighborhood density and phonotactic probability on word recognition, Vitevitch and Luce (1998) proposed that the effects take place at different levels of processing: neighborhood effects are due to competition at the lexical level and phonotactic effects are due to facilitation at a prelexical level (see also Luce and Large 2001 and Vitevitch and Luce 1999). However, both lexical neighborhood density and phonotactic probability reflect the similarity between the target word and other words in the language and the effects are in opposite directions, just like the opposite effects of near and distant semantic neighbors. Considering the definitions of lexical neighbors (difference of a single phoneme) and phonotactic probability (frequency of individual phoneme pairs), it is possible that lexical neighborhoods capture near phonological neighbor density and phonotactic probability captures distant phonological neighbor density. In this case, insights from semantic processing can shed light on phonological processing: the same type of contrasting patterns are captured by attractor-based models in the case of the homophony/polysemy effects and semantic neighborhood density effects, suggesting that if phonological processing were cast in attractor dynamics terms, both effects might emerge without being assigned to different levels of processing.

3 New Insights from Speech Perception

The preceding sections described two parallels between semantic ambiguity resolution and speech perception and emphasized the implications of these parallels on cognitive computational mechanisms and architecture. The following sections focus on three key issues in speech perception: (1) indirect effects and their role in the interactivity debate, (2) word frequency effects, and (3) illusions and noise and their implications for understanding optimal processing in the domain of language processing. Developments in understanding these issues in speech perception provide promising avenues for investigating semantic ambiguity resolution.

3.1 Indirect Effects and Interactive Processing

Most contextual effects are consistent with both interactive and autonomous models. For example, both TRACE and Merge predict that phoneme recognition should be faster in words than nonwords (Norris et al. 2000). As a result, these contextual effects cannot provide evidence to distinguish between the different models. The critical evidence for interactive processing in speech perception comes from a class of effects demonstrating prelexical consequences of lexical feedback (see McClelland et al. 2006 for a review).

Interactive models specifically predict that lexical information feeds back directly into prelexical processing so it should influence subsequent prelexical processes. Autonomous models integrate lexical and prelexical information at a decision stage, so it is impossible for lexical effects to have prelexical consequences.

The interactive-autonomous debate has focused on two prelexical effects that can be triggered by lexical feedback. The first demonstration was that lexical feedback could trigger effects of local auditory context (Elman and McClelland 1988; Magnuson et al. 2003b; Samuel and Pitt 2003). Local auditory context has been shown to influence speech perception: for example, an ambiguous sound between /k/ and /t/ will be more likely to be heard as /k/ if it is preceded by /s/ and as /t/ if it is preceded by /ʃ/ (Mann and Repp 1981). Lexical feedback has been shown to trigger this effect by replacing the precursor sound with an ambiguous one ('?' between /s/ and /ʃ/) and placing it in a disambiguating lexical context. For example, listeners were more likely to hear the ambiguous stop consonant as /k/ if it was preceded by *Christma?* (because the '?' was heard as /s/ to make the word *Christimas*) and as /t/ if it was preceded by *fooli?* (because the '?' was heard as /ʃ/ to make the word *foolish*).

A second important finding was that lexical feedback could trigger tuning or recalibration of speech perception (Eisner and McQueen 2005; Kraljic and Samuel 2005, 2006; Norris et al. 2003). In these studies an acoustically ambiguous speech sound (for example, one perceptually half-way between /s/ and /f/) was repeatedly presented in lexical contexts that were always biased toward a particular interpretation (for example, all context words were /s/-biased). Subsequent perception of the ambiguous phoneme was found to be shifted toward the lexically-consistent interpretation (in this example, /s/) even when the sound was presented in isolation. That is, lexical feedback informed earlier levels in the perceptual system that a particular sound— one that was previously ambiguous—corresponded to a particular phoneme and the system adjusted the mapping so that this sound was subsequently interpreted as that particular phoneme.

These effects have been controversial and proponents of the autonomous view have offered a number of possible non-interactive accounts of these data (e.g., McQueen 2003; though Magnuson et al. 2003a offer counterarguments to those accounts). Furthermore, as reviewed by McClelland et al. (2006) (see also discussion in McQueen et al. 2006 and Mirman et al. 2006b), currently the autonomous alternatives cannot capture the full pattern of data and the evidence for the interactive account of indirect effects continues to grow (Van Linden et al. 2007). Proponents of the autonomous view concede that lexically-guided tuning of speech perception requires feedback (Norris et al. 2003), thus this particular indirect effect has been particularly important. Lexically-guided tuning (like the other indirect effects) is a natural consequence of interactive processing (Mirman et al. 2006a, b; McClelland et al. 2006), though proponents of the autonomous view restrict the feedback required for this effect to be for learning only (Norris et al. 2003; McQueen et al. 2006).

Audio–visual speech perception (i.e., speech reading) is another domain in which the interactivity debate is highly relevant and indirect effects have been demonstrated. In the case of audio–visual speech perception there are autonomous models that integrate auditory and visual information at the decision level (e.g., the Fuzzy Logical Model of Perception Massaro 1998), models in which auditory and visual processing

are interactive in virtue of specifying the same underlying articulatory gestures (Fowler and Dekle 1991), and models in which interactivity develops as a result of experience with auditory–visual correlations (Stephens 2006). Just as lexical information can guide recalibration of speech perception, visual information can guide recalibration of auditory speech perception (Bertelson et al. 2003; Van Linden et al. 2007; Van Linden and Vroomen 2007). If an ambiguous speech sound (for example, between /b/ and /d/) is repeatedly presented in the context of disambiguating visual information (for example, the bi-labial closure of /b/), subsequent perception of the critical speech sound will be shifted to the visually-consistent interpretation even when the sound is presented in isolation. Like lexically-guided tuning, visually-guided tuning is an indirect effect that provides strong evidence for bi-directional interactions, in this case audio–visual interaction, rather than decision-level integration. Recent studies suggest that novel audio–visual correlations that are not based on articulatory gestures also produce audio–visual interaction effects (Stephens 2006), suggesting that interactive processing is at least partly a response to the structure of the input.

The interactive view of semantic ambiguity resolution makes the unique prediction that indirect effects of feedback should arise from semantic feedback as well as lexical feedback. For example, consider an ambiguous phoneme ('?', between /s/ and /ʃ/) in an ambiguous lexical context (for example, *accompli?*, which could be *accomplice* or *accomplish*) that is disambiguated by sentence context (for example, 'You should get help because this task requires an *accompli?*' versus 'You should get help because this task is difficult to *accompli?*'). Under the interactive view, sentence context directly influences lexical processing, which directly influences phoneme processing, thus the interactive view predicts that this sentence context manipulation should have lower-level consequences such as local auditory context effects and perceptual learning. A specific prediction of the effect of auditory context is that if the following sound is ambiguous between /k/ and /t/, perception should be biased towards /k/ in the *accomplice* case and /t/ in the *accomplish* case. A specific prediction of perceptual learning is that repeated presentation of the ambiguous sound in context should cause recalibration such that the ambiguous sound will be perceived in the contextually-appropriate way even when presented in isolation. Under the autonomous view, context effects are restricted to decision-level processes, thus the manipulation of sentence context should have no lower-level consequences (with the possible exception of perceptual learning under the feedback-for-learning-only account). It is important to note that such indirect effects will necessarily be small and difficult to demonstrate; nonetheless, evidence of such effects would make the strongest case for interactive models of semantic ambiguity resolution.

3.2 Frequency Effects

Ambiguous words tend to activate their more frequent meaning more than their less frequent meaning (Simpson and Burgess 1985). This result is just one example of many findings showing word frequency effects on word recognition and processing. The basic advantage for high frequency words (or meanings in the case of semantic ambiguity resolution) is consistent with at least three different mechanisms: (1) higher
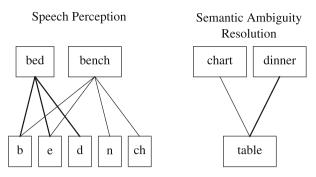
Speech Perception   Semantic Ambiguity
                                     Resolution

Fig. 2 Schematic diagram of frequency implementations in speech perception (*left*) and semantic ambiguity resolution (*right*) based on connection weights. Thicker lines indicate stronger connections. On the left, the phoneme /b/ is more strongly connected to the higher frequency word *bed* than to the lower frequency word *bench*. On the right, the word *table* is more strongly connected to the higher frequency meaning having to do with dinner than to the lower frequency meaning roughly synonymous with *chart*

frequency words have higher resting activation levels, (2) higher frequency words are preferred at a post-perceptual decision (or meaning selection) stage (sometimes called *ordered access*), or (3) higher frequency words have stronger bottom-up connections since they have been used more (this mechanism is illustrated in Fig. 2, where thicker lines indicate stronger connections). All three mechanisms predict an advantage for high frequency words, but they predict somewhat different time courses for this effect.

Dahan et al. (2001) implemented all three mechanisms of frequency effects in the TRACE model of speech perception (McClelland and Elman 1986) and provided concrete simulation demonstrations of the different predictions. Resting activation and decision level mechanisms both predicted strong early frequency effects, that is, high frequency words should have an early advantage that remains throughout the time course of processing. In contrast, the bottom-up connection strength mechanism predicted a gradually emerging high-frequency advantage, that is, initially there should be no difference between high and low frequency words and high frequency words should gradually build up an advantage. These different predictions for the time course of word frequency effects on recognition of unambiguous spoken words were tested using eye-tracking in the visual world paradigm (Dahan et al. 2001; Magnuson et al. 2007). In the visual world paradigm (VWP), subjects see several objects and hear verbal instructions to click on one of the objects. Subjects' eye fixations are closely time-locked to the spoken instructions (Tanenhaus et al. 1995; Allopenna et al. 1998; Magnuson et al. 2003c) and reveal the underlying processing at a finer grain than provided by traditional reaction time measures. Fixation probability during a VWP trial is taken as a continuous estimate of underlying lexical activation; by comparison, a priming effect is taken as a discrete measure of underlying lexical activation at a particular time point. For the case of word frequency, the influence of word frequency on word activation is reflected by greater fixation probability for high frequency words than low frequency words. The results (Dahan et al. 2001; Magnuson et al. 2007) showed that word frequency effects on spoken word recognition (i.e., differences in fixation probability) emerged early and gradually, as predicted by the bottom-up con-

nection strength account and conflicting with the resting activation and decision level accounts.

Models that learn the mapping from form to meaning (Kawamoto 1993; Plaut et al. 1996) naturally exhibit bottom-up connection-based frequency biases. Conceptually, for semantic ambiguity resolution, a bottom-up connection account of meaning frequency holds that there is a stronger association (connection) between the ambiguous word form and the more frequent meaning than the ambiguous word form and the less frequent meaning. Furthermore, this difference in connection strength is a monotonic function of the frequency difference; that is, a small difference in meaning frequency will correspond to a small difference in connection strength and a large difference in meaning frequency will correspond to a large difference in connection strength. Preliminary results suggest that meaning frequency effects also emerge early and increase gradually, as predicted by bottom-up connection strength accounts of frequency effects (Mirman et al. 2007). That is, initially there is no difference in activation of high and low frequency meanings, but high frequency meaning become more active more quickly than low frequency meaning.

The bottom-up connection strength view of frequency effects (both word and meaning frequency) naturally fits into a framework of integration of multiple graded constraints. This allows for graded and continuous trade-offs between, for example, perceptual effects, frequency effects, and context effects. Considering frequency and context effects in terms of continuous graded constraints is an important step forward from debates about discrete order-of-access for different meanings.

3.3 Noise, Illusions, and Optimality

Rational analysis (Anderson 1990) of cognitive architecture typically relies on an appeal to the notion of optimal processing. However, it is important to distinguish theoretically optimal processing from optimal processing in the typically noisy, yet highly constrained, context of real-world communication. Two issues—robustness to noise and illusory perceptions—have played an important role in understanding optimal processing in the context of speech perception.

Norris et al. (2000) argued that interactive feedback is sub-optimal because it does not improve word recognition even in deliberately interactive models such as TRACE. However, simulations of the TRACE model showed that feedback improves both speed and accuracy of word recognition (Magnuson et al. 2005). This effect was moderately beneficial under ideal input conditions (approximately 75% of all words in a large lexicon were recognized more quickly with feedback than without feedback) and became stronger when noise was added to the input (also, words that did not show a speed benefit of feedback showed an accuracy benefit under noise). This is an important finding because noise is an intrinsic aspect of cognitive processing (McClelland 1993) and a ubiquitous quality of the real-world communication signal. These simulations showed that given highly variable input, bi-directional connections help an interactive model (TRACE) separate the signal from the noise. It is possible to reap these benefits in an independent integration system as well; however, these benefits would be limited

to the decision stage and would not impact other aspects of processing (i.e., would fail to capture indirect effects as discussed in Sect. 3.1).

The TRACE simulations examined the benefits of word-to-phoneme feedback for word recognition, but comprehension is the goal of language processing. Semantic knowledge appears to feed back to phoneme processing. For example, an ambiguous utterance between *coat* and *goat* was more likely to be heard as *coat* in the context of a sentence such as 'The elderly grandma hurried to button the…' and as *goat* in the context of sentence such as 'The busy dairyman hurried to milk the…' (Borsky et al. 1998). It is reasonable to hypothesize that semantic feedback also helps word recognition; indeed, the benefits are likely to be even greater at the level of text or discourse comprehension because online interactions between bottom-up and top-down sources of information can radically constrain the possible interpretations of a word and thus make word comprehension and integration with the sentence and/or discourse much more efficient. Just as lexical (Magnuson et al. 2005) and visual (Ross et al. 2007) information helps the system deal with input noise, it is likely that discourse context information improves recognition of noisy inputs. By analogy to the spoken word recognition work, testing this hypothesis requires the implementation of a relatively large-scale interactive model of language processing and then examining the effect of manipulating the strength of feedback and amount of noise on word and/or sentence recognition performance.

Traditional modular or autonomous accounts argue that optimal processing requires a veridical representation of the perceptual input in order to avoid detrimental effects of context. In the domain of speech perception, this argument has generally taken the form of concerns over hallucination: if lexical information fed back to phoneme processing, then listeners might hallucinate lexically-consistent phonemes that were not present in the input. Proponents of the autonomous view proposed the possibility of illusory perceptions as a logical argument against interactive processing. However, the empirical fact is that behavioral data in multiple paradigms and domains demonstrate just such illusory effects of context. Listeners often fail to detect mispronunciations (Cole 1973; Marslen-Wilson and Welsh 1978), listeners report hearing lexically-consistent phonemes that were removed and replaced by noise (Samuel 1997; Warren 1970), and listeners are slower to recognize phonemes that are not consistent with the lexical context (Mirman et al. 2005). Such illusory, but contextually appropriate, perceptions are classic findings in a broad range of domains from illusory visual contours (Kanizsa 1979; Lee and Nguyen 2001) to false memory (Roediger and McDermott 1995; Sommers and Lewis 1999).

These detrimental effects illustrate that optimality must be defined specifically with respect to system environment and experience. Proponents of autonomous architectures equate optimality with veridical, error-free perception. However, in the domain of speech perception, there is tremendous pressure for rapid processing and the vast majority of speech input consists of known words, thus, using top-down lexical feedback to resolve perceptual ambiguity is useful in the vast majority of cases. Illusory perceptions are an inevitable consequence of processing speech input that violates these prior probabilities. A more flexible notion of optimality takes into consideration both the benefits (faster, more noise-resistant perception) and the costs (context-induced errors) of interactive processing. Interactive models naturally

balance the benefits and costs of context effects because feedback connections represent the prior probability of occurrence of perceptual units (Movellan and McClelland 2001). Autonomous models that integrate perceptual and contextual information at a decision stage could be made to fit the detrimental effects of context; however, such models would still fail to fit the indirect consequences of contextually-consistent perceptions (e.g., selective adaptation induced by lexically restored phonemes, Samuel 1997).

The critical point is that the optimal outcome is intrinsically dependent on the structure of the domain and may not coincide with intuitions, nor do intuitions always match behavioral findings. In addition, intuitions about optimality are themselves unstable: for speech perception appeals to optimality typically emphasize veridical perception of the input but for semantic ambiguity resolution the emphasis is typically placed on context-consistent interpretation of ambiguous words—exactly the type of outcome rejected as 'hallucination' for speech perception. Real-world language communication is very noisy but also very highly constrained: the vast majority of input is known words ordered in familiar structures, though occasionally new words and new meanings do occur. As a result, the language perceiver faces the classic *exploitation-exploration* dilemma: to exploit existing knowledge and context and thus risk missing new words or meanings or to preserve veridical input and risk interpreting noisy known words as novel words. Computational solutions to the exploitation-exploration problem, as well as behavioral and neural investigations, continue to be an active field of research (Daw et al. 2006). In sum, it is not trivial to determine the architectural implications of optimal language processing, nor is it parsimonious to argue for veridical perception in the case of speech perception and context-dependence in the case of semantic ambiguity resolution. Computational investigations provide a powerful tool for examining the structure of a domain and finding optimal solutions. So far, computational approaches have been mostly theory-driven: a researcher constructs a computational model based on a theory to test it. An alternative, bottom-up approach is to let a generic model develop a solution to the problem under consideration and examine the ways in which the result is consistent or inconsistent with theory. Both of these complementary approaches can be effective in furthering our understanding of the language processing system.

4 Concluding Remarks

The domains of speech perception and semantic ambiguity resolution share important structural similarities and the quest for general principles of language and cognitive processing calls for a consistent approach to solving both problems. Research at each level of language processing has typically been independent of research at other levels, however, sharing insights across levels has the potential to improve research in each. Recent developments in speech perception research point to promising avenues for research on semantic ambiguity resolution. First, the critical findings for the interactivity debate have come from indirect effects—consequences of feedback that cannot be due to post-perceptual information integration. Such findings, though controversial and potentially difficult to demonstrate, provide the strongest test of the interactive

view of both speech perception and semantic ambiguity resolution. Second, major advances in understanding the mechanism of word frequency effects have resulted from the use of eye-tracking in combination with computational modeling of proposed mechanisms. Similar approaches to investigating meaning frequency effects can resolve debates in semantic ambiguity resolution and provide significant steps toward a coherent theory of prior probability and exposure effects in language processing. Third, the concept of optimal processing, which has played a central role in theoretical debates on speech perception, must be considered in the context of the complexities (e.g., noise) and constraints (e.g., high probability of known and expected word) of real-world communication. These three issues reflect current advances in speech perception research and point to possible directions for advances in research on semantic ambiguity resolution. Sharing insights across levels of language processing is an important step toward developing a unified set of principles for language and cognitive processing.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory & Language, 38*(4), 419–439.

Anderson, J. (1990). *The adaptive character of thought*. Hillsdale, NJ: Erlbaum.

Bertelson, P., Vroomen, J., & de Gelder, B. (2003). Visual recalibration of auditory speech identification: A McGurk after effect. *Psychological Science, 14*(6), 592–597.

Borsky, S., Tuller, B., & Shapiro, L. P. (1998). "How to milk a coat": The effects of semantic and acoustic information on phoneme categorization. *Journal of the Acoustical Society of America, 103*(5), 2670–2676.

Buchanan, L., Westbury, C., & Burgess, C. (2001). Characterizing semantic space: Neighborhood effects in word recognition. *Psychonomic Bulletin & Review, 8*(3), 531–544.

Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics, 1*, 153–156.

Coltheart, M., Davelaar, E., Jonasson, J., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance VI* (pp. 535–555). Hillsdale, NJ: Erlbaum.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: Evidence from eye movements. *Cognitive Psychology, 42*(4), 317–367.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature, 441*(7095), 876–879.

Duffy, S. A., Morris, R. K., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*(4), 429–446.

Eisner, F., & McQueen, J. M. (2005). The specificity of perceptual learning in speech processing. *Perception & Psychophysics, 67*(2), 224–238.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*(2), 179–211.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory & Language, 27*(2), 143–165.

Fodor, J. (1983). *Modularity of mind*. Cambridge, MA: MIT Press.

Fowler, C. A. (2006). Compensation for coarticulation reflects gesture perception, not spectral contrast. *Perception & Psychophysics, 68*(2), 161–177.

Fowler, C. A., & Dekle, D. J. (1991). Listening with eye and hand: Cross-modal contributions to speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 17*(3), 816–828.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10*(4), 526–540.

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception & Performance, 16*(1), 77–91.

Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory & Language, 29*(1), 181–200.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception & Performance, 6*(1), 110–125.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*(4), 267–283.

Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance, 22*(6), 1331–1356.

Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research, 167*(1–2), 156–169.

Kanizsa, G. (1979). *Organization in vision*. New York: Praeger.

Kawamoto, A. H. (1993). Nonlinear dynamics in the resolution of lexical ambiguity: A parallel distributed processing account. *Journal of Memory and Language, 32*(4), 474–516.

Kraljic, T., & Samuel, A. G. (2005). Perceptual learning for speech: Is there a return to normal? *Cognitive Psychology, 51*(2), 141–178.

Kraljic, T., & Samuel, A. G. (2006). Generalization in perceptual learning for speech. *Psychonomic Bulletin & Review, 13*, 262–268.

Lee, T. S., & Nguyen, M. (2001). Dynamics of subjective contour formation in early visual cortex. *Proceedings of the National Academy of Sciences, 98*(4), 1907–1977.

Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes, 16*(5/6), 565–581.

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

MacDonald, M. C., & Pearlmutter, N. J., & Seidenberg, M. S. (1994). The lexical nature of syntactic ambiguity resolution. *Psychological Review, 101*(4), 676–703.

Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 1–24.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003a). Lexical effects on compensation for coarticulation: A tale of two systems? *Cognitive Science, 27*(5), 801–805.

Magnuson, J. S., McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2003b). Lexical effects on compensation for coarticulation: The ghost of Christmash past. *Cognitive Science, 27*(2), 285–298.

Magnuson, J. S., Mirman, D., & Harris, H. D. (2008). Computational models of spoken word recognition. In M. Spivey, M. Joanisse, & K. McRae (Eds.), *The Cambridge handbook of psycholinguistics*. Cambridge: Cambridge University Press.

Magnuson, J. S., Strauss, T., & Harris, H. D. (2005). Interaction in spoken word recognition models: Feedback helps. In B. G. Bara, L. Barsalou, & M. Bucciarelli (Eds.), *Proceedings of the 27th annual conference of the cognitive science society* (pp. 1379–1384). Mahwah, NJ: Lawrence Erlbaum Associates.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., & Dahan, D. (2003c). The time course of spoken word learning and recognition: Studies with artificial lexicons. *Journal of Experimental Psychology: General, 132*(2), 202–227.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America, 69*, 546–558.

Marr, D. (1982). *Vision: A computational approach*. San Francisco, CA: W. H. Freeman.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*(1), 29–63.

Massaro, D. W. (1998). *Perceiving talking faces: From speech perception to a behavioral principle*. Cambridge, MA, USA: The MIT Press.

McClelland, J. L. (1993). Toward a theory of information processing in graded, random, interactive networks. In D. E. Meyer & S. Kornblum (Eds.), *Attention & performance XIV: Synergies in experimental*

psychology, artificial intelligence and cognitive neuroscience (pp. 655–688). Cambridge, MA: MIT Press.

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

McClelland, J. L., Mirman, D., & Holt, L. L. (2006). Are there interactive processes in speech perception? *Trends in Cognitive Sciences, 10*(8), 363–369.

McQueen, J. M. (2003). The ghost of Christmas future: didn't Scrooge learn to be good? Commentary on Magnuson, McMurray, Tanenhaus, and Aslin (2003). *Cognitive Science, 27*(5), 795–799.

McQueen, J. M., Norris, D., & Cutler, A. (2006). Are there really interactive processes in speech perception? *Trends In Cognitive Sciences, 10*(12), 533.

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, and Cognition, 34*(1), 65–79.

Mirman, D., Strauss, T., Magnuson, J. S., & Dixon, J. A. (2007, March). *Integration of pragmatic context in homophone ambiguity resolution: Time course of activation of context-appropriate and context-inappropriate meanings*. Poster presented at the 20th Annual CUNY Conference on Human Sentence Processing, La Jolla, CA.

Mirman, D., McClelland, J. L., & Holt, L. L. (2005). Computational and behavioral investigations of lexically induced delays in phoneme recognition. *Journal of Memory & Language, 52*(3), 424–443.

Mirman, D., McClelland, J. L., & Holt, L. L. (2006a). An interactive Hebbian account of lexically guided tuning of speech perception. *Psychonomic Bulletin & Review, 13*(6), 958–965.

Mirman, D., McClelland, J. L., & Holt, L. L. (2006b). Reply to McQueen et al.: Theoretical and empirical arguments support interactive processing. *Trends in Cognitive Sciences, 10*(12), 534.

Mirman, D., McClelland, J. L., Holt, L. L., & Magnuson, J. S. (2008). Effects of attention on the strength of lexical influences on speech perception: Behavioral experiments and computational mechanisms. *Cognitive Science* (in press).

Movellan, J. R., & McClelland, J. L. (2001). The Morton-Massaro law of information integration: Implications for models of perception. *Psychological Review, 108*(1), 113–148.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral & Brain Sciences, 23*(3), 299–370.

Norris, D., McQueen, J. M., & Cutler, A. (2003). Perceptual learning in speech. *Cognitive Psychology, 47*(2), 204–238.

O'Craven, K., Rosen, B., Kwong, K., Treisman, A., & Savoy, R. (1997). Voluntary attention modulates fMRI activity in human MT-MST. *Neuron, 18*(4), 591–598.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception & Performance, 19*(4), 699–725.

Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology, 29*(2), 149–188.

Pitt, M. A., & Samuel, A. G. (2006). Word length and lexical activation: Longer is better. *Journal of Experimental Psychology: Human Perception and Performance, 32*(5), 1120–1135.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review, 103*(1), 56–115.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*(3), 372–422.

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

Rodd, J. M., Gaskell, M. G., & Marslen-Wilson, W. D. (2004). Modelling the effects of semantic ambiguity in word recognition. *Cognitive Science, 28*(1), 89–104.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(4), 803–814.

Ross, P., Saint-Amour, D., Leavitt, V., Javitt, D., & Foxe, J. (2007). Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral Cortex, 17*(5), 1147–1153.

Rubin, P., Turvey, M. T., & Van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in spoken nonwords. *Perception & Psychophysics, 19*(5), 394–398.

Samuel, A. G. (1997). Lexical activation produces potent phonemic percepts. *Cognitive Psychology, 32*(2), 97–127.

Samuel, A. G., & Pitt, M. A. (2003). Lexical activation (and other factors) can mediate compensation for coarticulation. *Journal of Memory & Language, 48*(2), 416–434.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982) Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14*(4), 489–537.

Simpson, G., & Burgess, C. (1985). Activation and selection processes in the recognition of ambiguous words. *Journal of Experimental Psychology: Human Perception and Performance, 11*(1), 28–39.

Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language, 40*(1), 83–108.

Stephens, J. (2006). *The role of learning in audiovisual speech perception*. Ph.D. dissertation, Carnegie Mellon University.

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*(5217), 632–634.

Twilley, L. C., & Dixon, P. (2000). Meaning resolution processes for words: A parallel independent model. *Psychonomic Bulletin & Review, 7*(1), 49–82.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics, 34*(5), 409–420.

Van Linden, S., Stekelberg, J., Tuomainen, J., & Vroomen, J. (2007). Lexical effects on auditory speech perception: An electrophysiological study. *Neuroscience Letters, 420*(1), 49–52.

Van Linden, S., & Vroomen, J. (2007). Recalibration of phonetic categories by lipread speech versus lexical information. *Journal of Experimental Psychology: Human Perception and Performance*, *33*(6), 1483–1494.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing in perception of spoken words. *Psychological Science, 9*(4), 325–329.

Vitevitch, M. S., & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory & Language, 40*(3), 374–408.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167*(3917), 392–393.