Listen, You Know What I Mean ...
– But How Do You Get There?

Lena Kästner

1 Introduction

A mystical story surrounds Pink Floyd's "Another Brick in the Wall" (Harder, 1996). Shortly after the song had been recorded Peter Fischer, the band's German sound engineer, disappeared and something strange was found in the recording: the children's choir clearly sang the German line "Holt ihn, holt ihn unters Dach!" ("Get him, get him under the roof!") in the chorus. All members of the choir assured to have sung the original English text "*All in all it's just a*nother brick in the wall", however. Fischer was found hanged in the studio's attic. Investigations revealed he had formerly worked in an orphanage and abused children. Was this their late revenge?

If you listen to the song carefully after reading this, you will easily make out the German line – but the whole story is pure fiction! This nicely illustrates how easily we can be deceived in our auditory perception. But why?

Speech processing is not just a matter of translating sounds into neural code; it requires the recognition of words, parsing of sentences, and inferring of meaning. And potential flaws lure at any step. Why can we still communicate effectively? – Basically, it is because we ignore irregularities, guess what is most likely to come next in a train of speech, and use multiple sources to disambiguate the perceived input.

This essay focuses on auditory speech perception in the adult brain. It will sketch the basic properties of the speech signal, a series of psycholinguistic or computational processing steps required to analyse it, and review empirical evidence to illuminate how auditory language comprehension is possibly realised in the brain. In due course, the challenges a listener faces will be considered. I will close with a number of questions subject to future research.

2 The Speech Signal: Trains of Phonemes

Speech is sound produced by vocal organs of a speaker; it typically ranges from 80Hz (deep male voice) to 400Hz (child voice). The signal is remarkably variable: there are different languages, dialects, and accents, different speakers with different voices, different paces and pitches of speech, different stresses, rhythms and prosody, . . . At the same time, the rate of data transmission is extremely high: *"I met a friend yesterday and we went to the British Museum."* takes only 2.7 seconds to utter but contains 12 (not even clearly separated) words, 17 syllables, 46 phonemes, and even prosodic information.

Phonemes are the smallest perceptually distinct units of speech. The International Phonetic Alphabet (IPA) classifies 100 different phonemes across languages. Not every language uses the full range, however. For instance, in English aspirated and non-aspirated "p" are *allophones*, i.e. their substitution will not cause a change in word meaning. But in Thai aspirated and unaspirated "paa" (meaning "spilt" and "forest", respectively) form a *minimal pair* (Ward, 2006).

Our sensitivity to non-native speech sound differences declines drastically during the first year of life (e.g. Werkera & Tees, 2002). This "native tuning" is useful e.g. because it reduces the number of possible input interpretations. It will thus be less problematic for a listener that phonemes are influenced by pre- and proceeding phonemes in normal speech as such variations are less likely to change word meaning; an analogous claim explains why we can understand accents.

Another effect of language experience is *categorical perception* (Ramez *et al.*, 2001): to most English speakers, e.g. "ba" and "pa" are clearly distinct though they actually vary continuously along a spectrum. Categorical perception helps us deal with different voices but it also may also cause problems of initial misunderstanding. For example, I may realise only during semantic processing, that the dog likely is not "parking" but "barking".

Even in a monolingual environment without noise or interfering voices, speech sounds are often too ambiguous to recognise their constituent phonemes. For effective communication, the listener will have to rely on additional cues – be it semantics (as with the barking dog), lip patterns, conversational context, grammar, general world knowledge, or even pragmatics – when processing the continuous unlabeled stream.

3 Word Recognition, Sentence Parsing & Interpretation

Psycholinguistic theories and computational models of auditory language processing employ phonemes as the basic building blocks fed into a computational chain. Once phonemes have successfully been identified, we can recognise words, access word meaning, construct sentences, and interpret them.

3.1 Words

Word recognition is an incremental process heavily affected by context and language experience. Warren (1970) demonstrated that listeners perceive a complete word even in absence of its last syllable. This *restoration effect* illustrates how "clever guesses" aid language comprehension. Marslen-Wilson (1987) suggested a *cohort model* according to which all words initially consistent with the input are processed in parallel until enough of the input is revealed to identify a single match. Recent eye-tracking studies support this view (e.g. Spivey *et al.*, 2002). Similarly, Hartmann (2004) found that German speakers identify the target stimulus even prior to its onset if preceded by an article indicating gender and/or an agreeing adjective. Semantic and/or syntactic processing may thus figure in word recognition questioning the sequentiality of speech processing so often assumed.

Computational models, e.g. Finite State Transducers or Markov Chains (see Jurafsky & Martin, 2009), highlight how guesses as to how an auditory stimulus continues can be based on the statistical properties of a language (word frequency, likelihood of a combination of phonemes, inflectional morphology) we learn through experience. Indeed, native language experience appears to play a pivotal role in word recognition, as Spivey and Marian (1999) demonstrated. Hahne (2001) made an analogous observation at the level of sentences. This evidence speaks in favour of “native tuning” beyond the level of phonemes.

The facilitating nature of context (including grammar and linguistic experience) is particularly useful where speech signals are incomplete, noisy, or highly ambiguous – but successful guesses are not guaranteed, otherwise we would be less easily deceived.

3.2 Sentences: Who does What to Whom?

Word recognition alone is insufficient to parse sentences and infer their meaning. For instance, “till” can have differential meanings depending whether it is used as noun or preposition. Similarly, “Mary hits John.” and “John hits Mary.” consist of the same words but have very different meanings. Words thus need to be tagged with categories and features (gender, case, number, time, transitivity, etc.) before syntactic and sentence-level semantic analysis can take place.

However, some semantic and/or syntactic interpretation may be required beforehand to disambiguate the identify a grammatically licensed word sequence. The exact order of operations is as much a matter of debate as how we achieve them.

Frazier and Rayner (1982) argue for *structure driven parsing*, i.e. that a sentence’s parse is determined solely by the syntactic features of its constituent words. Computational models such as *recursive descent* and *shift-reduce* parsers illustrate the difficulties with mere structure driven approaches: expensive backtracking and re-analysis are often required (cf. Jurafsky & Martin, 2009). However, the view is supported by *garden path sentences* like "Since John jogs a mile seems a short distance to him." Here, the common parse will not be correct as there is a parsing preference for noun-phrase-plus-verb over noun-phrase-plus-relative-clause. This preference for particular structures, again, likely is a matter of (native) language experience. Some parsing algorithms, e.g. the probabilistic CKY (Cocke-Kasami-Younger-Kasami) algorithm (Ney, 1991), take this into account and rely on frequencies associated with particular structures for disambiguation. Empirical studies (e.g. Trueswell, Tanenhaus & Kello, 1994) indicate similar principles may underlie parsing in the brain.

If semantic factors (word meaning, context) aid parsing, it could be *discourse driven* (MacDonald, Pearlmutter, & Seidenberg, 1994). This view is supporedt by experiments demonstrating how visual context helps resolve parsing ambiguities (Altmann *et al.*, 1994; Eberhard *et al.*, 1995). Similarly, *prosody* may act as syntactic markers and aid parsing (Friederici, Meyer & von Cramon, 2000).

What cues and strategies the brain actually employs to analyse sentences is unclear. Recently, models of language processing as probabilistic inference have come into focus (e.g. Chater & Manning, 2006); hopefully, these will illuminate how the brain processes speech.

3.3 Intended Meaning

Once a sentence has successfully been parsed, the listener can infer its meaning. This is not necessarily as straightforward as assessing semantic knowledge of the words and their relations. Gestures, body language, prosody, . . . may significantly affect conversational meaning as utterances may be ironic, or may have implications beyond literal meaning.

To deal with this, Grice (1975) suggested speakers should make their conversational contribution *qualitatively* and *quantitatively adequate* (no lies, not telling too little/much), *relevant*, and in a *non-obscuring* and *cooperative* fashion. These maxims taken for granted, the intended or *pragmatic* meaning of a proposition may deviate from its literal interpretation: if you brought two bottles of wine and are asked whether you brought "a bottle", "No." – the logically correct answer – will probably not be considered appropriate.

Pragmatics may be considered as *reasoning* rather than speech processing and receives little attention in speech processing brain research. Note though, that pragmatic considerations need not be restricted to a final processing stage.

4 A Neural Basis for Auditory Sentence Processing

4.1 From the Air to A1

To process auditory speech signals, the brain must first translate them into neural code. When vibrating air molecules hit our external ears, sound is selectively amplified and directed towards the middle ear. The signal propagates through the tympanic membrane and ear ossicles which mechanically transmit sound waves from air to the fluid inside the cochlea causing vibrations of the basilar membrane; these are translated into neural signals and sent down the auditory nerve to primary auditory cortex (A1). If this system is not finely tuned, perception may in accurate.

A1 lies on the medial surface of the temporal lobe bilaterally in *Heschel's gyri*; it is tonotopically organised, containing complex spectotemporal receptive fields (Talavage *et al.*, 2004; Wang *et al.*, 2005). Temporal and frequency information is encoded *via* different pathways projecting to A1. At this early stage, speech processing is not lateralised (Binder *et al.*, 2000) although language is typically associated with a left-lateralised perisylvian network (Friederici, Meyer, & von Cramon, 2000; MacSweeney *et al.*, 2008; Price, Thierry, & Griffiths, 2005).

4.2 The Left Brain's "What" and "How"

The left-lateralisation of language has initially been emphasised by two types of aphasia both due to damage in left-hemispheric regions. Patients with *Broca's aphasia* (1861) typically suffer from articulatory deficits and agrammatism; they have difficulties discriminating syllables while auditory comprehension was preserved (Blumstein *et al.*, 1977). Patients with *Wernicke's aphasia* (1874) show the reverse pattern, i.e. fluent but non-sentential speech, poor auditory comprehension, but intact syllable discrimination (Ward, 2006). It has thus been hypothesised that damage to Broca's area (BA44/45) affects speech motor output whereas damage to Wenicke's area (posterior BA22) affects speech motor comprehension (Geschwind, 1967).

Finding that the two syndromes are less distinct than initially assumed (e.g. Caramazza & Zurif, 1976), challenged this view. Still, contemporary theories of speech perception build on them suggesting, e.g., that there are at least two functionally dissociable left-lateralised routes of speech perception from A1 onwards. Hickok and Poeppel (2000) argue that sound based representations feed, in a task-dependent manner, either into a ventral *what* stream connecting to distributed conceptual systems *via* the tempo-parietal-occipital junction or into a dorsal *how* stream *via* inferior parietal, and frontal areas. Compatible with this view, Ward (2006) includes the superior temporal sulcus (STS), superior temporal gyrus (STG), temporal pole (associated with *semantic knowledge*; Patterson, Nestor & Rogers, 2007), and Broca's area into the lexico-semantic *what* route while Wernicke's area, the angular gyrus and, again, Broca's area constitute the *how* stream. Whereas the lexico-semantic route is recruited for tasks requiring access to the *mental lexicon*, the *how* stream is recruited for access to sub-lexical segments, e.g., phoneme recognition which may be based on their corresponding motor commands (Liberman & Wahlen, 2000). The proposed distinction between *what* and *how* streams of speech perception parallels that between visual *what* and *where* streams (Ungerleider & Mishkin, 1982) and is supported by studies in auditory object identification (Scott & Johnsrude, 2003). The hypothesised double-role for Broca's area may be due to its involvement in working memory and monitoring processes (see section 4.3).

An initial phonetic analysis of the speech signal may be localised predominantly in superior temporal regions as speech processing typically aims at conveying content and will thus involve the lexico-semantic route wherein phoneme perception precedes accessing semantic stores. On-line analysis of an input's specto-temporal structure (silent gaps, noise burst, formant transitions) crucial for classification of speech sounds has been found to implicate STS and STG (Leech *et al.*, 2009). Phoneme recognition in these areas may be aided by top-down connections from higher-level areas accounting for categorical perception (Binder *et al.*, 2004). This begs the question of relevance for the *how* route previously described.

4.3 Syntax and Semantics

Taking into account the considerations in section 3, it seems speech processing may not be a (fully) sequential process. Investigations into its neural basis are thus complicated as processes occurring all at once and/or being carried out in the same structures are difficult to separate. Sequential processing steps are more easily investigated and tempting to assume. Indeed, fractioning of cortical regions into increasingly complex processing areas has been suggested (Davies & Johnsrude, 2003). Even if the *involvement* of higher brain areas in theoretically higher-level processing can be verified with experimental procedures, this does not guarantee such areas are actually triggered by linguistic processing. For instance, Price *et al.* (2005) demonstrated that any part of "speech-specific" STS is at least recruited by either environmental sounds, melody, or conceptual processing.

Although semantics may interact with syntax, a dissociation between syntactic and semantic processing is supported neuropsychological cases: patients with *semantic dementia* are selectively impaired on semantic tasks but do not display syntactic impairments (e.g. Warrington & Shallice, 1984) while Ostrin & Tyler (1995) report a patient with the reverse pattern.

Assuming syntax and semantics are indeed neurally dissociable, Friederici, Meyer & von Cramon (2000) suggest syntax to be primarily processed in left posterior STG, bilateral STS, and left inferior frontal cortex while part of Broca's area (pars triangularis), the inferior frontal gyrus bilaterally, left STS, and the banks of posterior STS are crucial for semantic processing. The involvement of frontal areas in both cases may be due to strategic and/or memory aspects in language processing where BA45/47 is relevant for the semantic, and BA44 for the syntactic domain (Friederici, 2002). Thus, syntactic and semantic routes could be subdivisions of the *what* stream. To what extent any of these areas are actually recruited by language *per se* rather than thematic processing, structure building, and general working memory, is debatable (Fiebach *et al.*, 2001; Newman *et al.*, 2003).

Whilst lesion and imaging data illuminate localisation, electrophysiology helps investigate the time course of speech processing. Some EPR components are associated with specific processing steps: a negativity peak 400ms after word onset, *N400*, is observed if a word cannot be semantically integrated (Kutas & Hillyard, 1984); *left-anterior negativity* (LAN) is associated with morphosyntactic errors such as incorrect inflection (Coulson, King & Kutas, 1998); an *early LAN* (ELAN) is observed in connection with word-category mistakes (Friederici, Pfeifer & Hahne, 1993); and a late (after 600ms) posterior positivity, *P600*, is associated with syntactic violations and garden-path sentences (Osterhout & Holcomb, 1992). What does this mean for the time course of speech processing?

One interpretation is that syntactic processing precedes semantic interpretation as ELAN occurs prior to N400 (Pulvermüller *et al.*, 2008). Building on the idea of syntax-driven parsing, Frazier and Fodor (1987) suggest a model of such a *syntax first view*. Marslen-Wilson and Tyler (1980), on the other hand, argue for an *interactive view* according to which all types of linguistic information interact at all times. Friederici (2000) combines both views; her *neurocognitve model* postulates three phases of auditory processing: an early (100-300ms) phase of parsing that relies on word-category information, an intermediate (300-500ms) phase of lexico-semantic and morphosyntactic processing leading to thematic role assignment, and a late (500-1000ms) information integration phase during which initial parses are revised (see Figure 1). Thus syntax is processed autonomously at first while at a later stage different types of information interact.

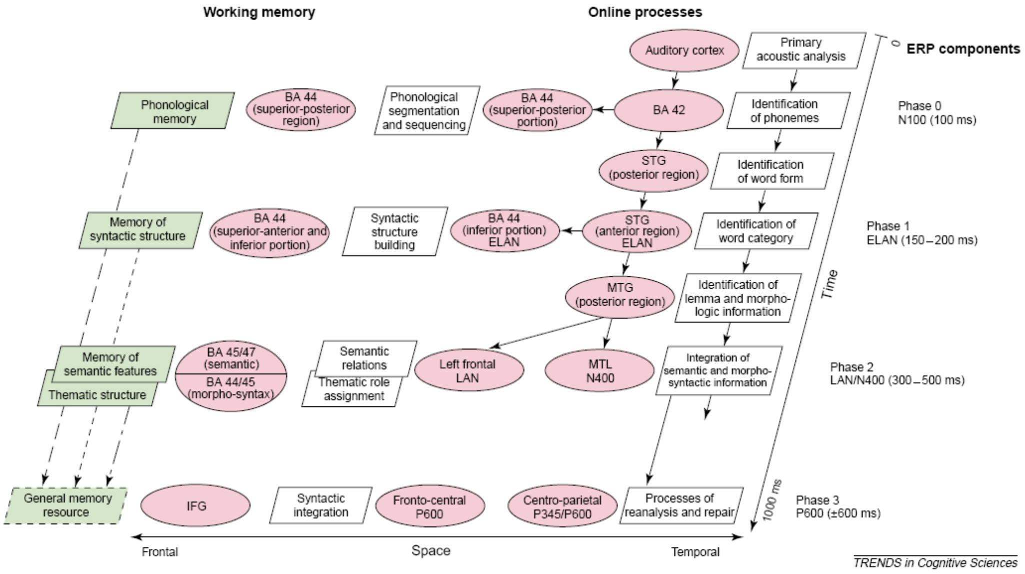

Figure 1: Adopted from Friederici (2002).

Friederici's model builds on electrophysiological data indicative of error processing in the brain (ELAN, LAN, N400, P600). It accounts for single errors and combinations of them: if word-category and semantic violations occur together, only an ELAN but no N400 is observed (Friederici, Steinhauer & Frisch, 1999). This is consistent with the model's prediction of syntax being processed initially: the non-occurrence of N400 indicates lexical integration to be prohibited while semantic processing takes place. Similarly, the interaction of different types of information in the late window is supported by observed P600 modulation when morphosyntactic and semantic violations are combined (Gunter, Friederici & Schiefers, 2000).

A possible extension for Friederici's model provides the finding that (global) phrase structure building and (local) verb argument processing – two theoretically different syntactic processes – differentially correspond to ERP patterns, viz. ELAN/P600 and N400/P600, respectively (Friederci & Meyer, 2004).

As a note of caution, it should be added that the model substantially builds on data interpretation in terms of ERP components. Some of these (and thus the model) may be found to be in need of revision if additional components of language processing, possibly interfering with known ones, are discovered.

5 Future Directions

In due course, challenges occurring during speech perception – ranging from phoneme recognition to pragmatic interpretation – have been outlined. As an attempt to understand how the brain possibly deals with these problems, *what* and *how* routes (see section 4.2) as well as a neurocognitive model of auditory sentence processing (see section 4.3) have been considered. Despite their explanatory strengths, both should be taken with a pinch of salt: the relevance of the *how* stream for higher-level processing and its integration with the lexico-semantic route are not entirely clear; and many "speech-specific" regions included in neural models are not *just* speech specific.

Assuming there are macro-anatomical modules neatly corresponding to psycholinguistic/computational processing steps is almost certainly flawed. A more biologically plausible suggestion is that speech recognition emerges from connectivity (Price *et al.*, 2005) between auditory areas (perhaps specifically tuned to the statistical properties of the auditory input we experience) that builds the basis for conceptual and acoustic integration over different time scales. But how exactly this integration is achieved remains unclear. Possibly, neural oscillations could be employed, as has been suggested at other occasions where cortically distributed information needs to be integrated (Bressler & Kelso, 2001; Buzsaki & Draguhn, 2004; Hart & Kraut, 2007; Tallon-Baudry, & Bertrand, 1999).

Likewise, semantic memory structure and retrieval should be considered more closely in future. Transforming auditory input into meaningful concepts is required when words are recognised, sentences parsed, and flawed interpretations rejected. How this transformation is achieved remains unclear to date (cf. Formisano *et al.*, 2008) but will likely depend on how semantic stores are organised.

Finally, cross-linguistic differences should be given consideration. Parsing involves building phrase structures consistent with perceived word order. Permissible word order varies with language and so do *surface structure building* and *thematic role assignment* (e.g. Carnie, 2002). If already two linguistically different operations (Friederci & Meyer, 2004) can neurally be distinguished, at least some aspects of speech processing may differ cross-linguistically.

In summary, we have seen that for effective communication the speech signal must be processed through a complex analysis. A listener will often have to rely on additional cues – be it semantics, lip patterns, conversational context, grammar, general world knowledge, or even pragmatics – to disambiguated the unlabeled sound waves. Native tuning, language experience, and top-down influences, e.g. from grammar in phoneme recognition, or from context in sentence recognition, likely have a significant impact. If you had not been told to listen for the German line in Pink Floyd's song, you would probably never have heard it.

References

Altmann, G.T.M., Garnham, A., & Henstra, J.-A. (1994). Effects of syntax parsing: evidence against a structure-based proposal mechanism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, 209-216.

Binder, J.R., Frost, J.A., Hammeke, T.A., Bellgowan, P.S., Springer, J.A., Kaufman, J.N., & Possing, E.T. (2000). Human temporal lobe activation by speech and nonspeech sounds. *Cerebral Cortex*, 10, 512-528.

Binder, J.R., Liebenthal, E., Possing, E.T., Medler, D.A., & Ward, D. (2004) Neural correlates of sensory and decision processes in auditory object identification. *Nature Neuroscience*, 7, 295-301.

Blumstein, S.E., Cooperc, W.E., Zurife, E.B., Caramazza, A. (1977). The perception and production of voice onset time in aphasia. *Neuropsychologia*, 15, 371-383.

Bressler, S.L. & and Kelso, J.A.S. (2001). Cortical coordination dynamics and cognition. *Trends in Cognitive Sciences*, 5, 26-36.

Broca, P. (1861). Remarques sur le siege de la faculte du langage articule suives d'une observation d'aphŐmie. *Bulletin et Memoires de la Societe Anatomique de Paris*, 2, 330-357.

Buzsaki, G. & Draguhn, A. (2004). Neuronal oscillations in cortical networks. *Science*, 304, 1926-1929.

Carnie, A. (2002). *Syntax: A Generative Introduction*. Oxford: Blackwell.

Chater, N. & Manning, C.D. (2006). Probabilistic models of language processing and acquisition- *Trends in Cognitive Sciences*, 10, 335-344.

Coulson, S., King, J.W., & Kutas, M. (1998). Expect the unexpected: event-related brain responses of morphosyntactic violations. *Language and Cognitive Processes*, 13, 21-58.

Davies, M.H. & Johnsrude, I.S. (2003). Hierarchical processing in spoken language comprehension. *Journal of Neuroscience*, 28, 3423-3431.

Eberhard, K.M., Spivey-Knowlton, M.J., Sedivy, J.C., & Tanenhaus, M.K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research*, 24, 409-436.

Fiebach, C.J., Schlesewsky, M., & Friederici, A.D. (2001). Syntactic working memory and the establishment of filler-gap dependencies: insights from ERPs and fMRI. *Journal of Psycholinguistic Research*, 30, 321-338.

Formisano, E., De Martino, F., Bonte, M., & Goebel, R. (2008). "Who" is saying "what"? Brain-based decoding of human voice and speech. *Science*, 322, 970-973.

Frazier, L. & Fodor, J.D. (1978). The sausage machine: a new two-stage model of the parser.*Cognition*, 6, 291-325.

Frazier, L. & Rayner, K. (1982). Making and correcting errors in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.

Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing. *Trends in Cognitive Sciences*, 6, 78-84.

Friederici, A.D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, 1, 183-192.

Friederici, A.D., Steinhauer, K., & Frisch, S. (1999) Lexical integration: sequential effects of syntactic and semantic information. *Memory and Cognition*, 27, 438-453.

Friederici, A.D. & Meyer, M. (2004). The brain knows the difference: two types of grammatical violations. *Brain Research*, 1000, 72-77.

Friederici, A.D., Meyer, M., & von Cramon, D.Y. (2000). Auditory language comprehension: an event-related fMRI study on the processing of syntactic and lexical information. *Brain and Language*, 74, 289-300.

Geschwind, N. (1967). The varieties of naming errors. *Cortex*, 3, 97-112.

Grice, P. (1975). Logic and Converstaion. In P. Cole & J. Morgan (Eds.), *Syntax and Semantics*. Reprinted in P. Grice, *Studies in the Way of Words* (pp. 22-40). Cambridge, MA: Harvard University Press, 1991.

Gunter, T.C., Friederici, A.D., & Schriefers, H. (2000). Syntactic gender and semantic expectancy: ERPs reveal early autonomy and late interaction. *Journal of Cognitive Neuroscience*, 12, 556-568.

Hahne, A. (2001). WhatÕs different in second-language processing? Evidence from event-related brain potentials. *Journal of Psycholinguistic Research*, 30, 251-266.

Harder, B. (1996). Die §bersinnlichen PhŁnomene im Test. Augsburg: Pattloch Verlag.

Hart, J.Jr., & Kraut, M.A. (2007). Neural hybrid model of semantic object memory (version 1.1). In J. Hart, Jr., & M.A. Kraut (Eds.), *Neural Basis of Semantic Memory* (pp. 331-359). Cambridge: Cambridge University Press.

Hartmann, N. (2004). *Processing Grammatical Gender in German – An Eye-Tracking Study on Spoken-Word Recognition*. (Dissertation, University of Osnabr§ck). Retrieved from http://cogsci.uni-osnabrueck.de/CL/download/Hartmann$_G$ramGender.pdf

Hickok, G. & Poeppel, D. (2000). Towards a functional neuroanatomy of speech perception. *Trends in Cognitive Sciences*, 4, 131-138.

Jurafsky, D. & Martin, J.H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Speech Recognition, and Computational Linguistics*. Upper Saddle River, NJ: Prentice-Hall.

Kutas, M. & Hillyard, S.A. (1984). Brain potentials during reading reflect word expectancy and semantic association. *Nature*, 307, 161-163.

Leech, R., Holt, L.L., Devlin, J.T., & Dick, F. (2009). Expertise with artificial nonspeech sounds recruits speech-sensitive cortical regions. *Journal of Neuroscience*, 29, 5234-5239.

Liberman, A.M. & Wahlen, D.H. (2000). On the relation of speech to language. *Trends in Cognitive Sciences*, 4, 187-196.

MacSweeney, M., Capek, C.M., Campbell, R., & Woll, B. (2008). The signing brain: the neurobiology of sign language. *Trends in Cognitive Sciences*, 12, 432-440.

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

Marslen-Wilson, W.D. & Tyler, L.K. (1980). The temporal structure of spoken language understanding. *Cognition*, 8, 1-71.

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

Ney, H. (1991). Dynamic programming parsing for context-free grammars in continuous speech recognition. *IEEE Transactions on Signal Processing*, 39, 336-340.

Osterhout, L. & Phillip J. Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, 785-806.

Ostrin, R.K. & Tyler, L.K. (1995). Dissociations of lexical function: semantics, syntax, and morphology. *Cognitive Neuropsychology*, 12, 345-389.

Patterson, K., Nestor, P.J., & Rogers, T.T. (2007). Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature Reviews Neuroscience*, 8, 976-987.

Price, C., Thierry, G., & Griffiths, T. (2005). Speech-specific auditory processing: where is it? *Trends in Cognitive Sciences*, 9, 271-276.

Pulvermüller, F., Shtyrov, Y., Hasting, A.S., & Carlyton, R.P. (2008). Syntax as a reflex: neurophysiological evidence for the early automaticity of syntactic processing. *Brain and Language*, 104, 244-253.

Remez, R.E., Pardo, J.S., Piorkowski, R.L., & Rubin, P.E. (2001). On the bistability of sine wave analogues of speech. *Psychological Science*, 12, 24-29.

Scott, S.K. & Johnsrude, I.S. (2003). The neuroanatomical and functional organization of speech perception. *Trends in Neurosciences*, 26, 100-107.

Talavage, T.M., Sereno, M.I., Melcher, J.R., Ledden, P.J., Rosen, B.R., & Dale, A.M. (2004). Tonotopic organization in human auditory cortex revealed by progressions of frequency sensitivity. *Journal of Neurophysiology*, 91, 1282-1296.

Tallon-Baudry, C. & Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. *Trends in Cognitive Sciences*, 3, 151-162.

Spivey, M.J., Tanenhaus, M.K., Eberhard, K.M., & Sedivy, J.C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology*, 45, 447-481.

Trueswell, J.C., Tanenhaus, M.K., Kello, C. (1993). Verb-specific constraints in sentence processing: separating effects of lexical preference from garden-paths. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 528-553.

Ungerleider, L.G. & Mishkin, M. (1982). Two cortical visual systems. In D.J. Ingle, M.A. Goodale, & R.J.W. Mansfield (Eds.), *Analysis of Visual Behavior* (pp. 549-586), Cambridge, MA: MIT Press.

Wang, X., Lu, T., Snider, R.K., & Liang, L. (2005). Sustained firing in auditory cortex evoked by preferred stimuli. *Nature*, 435, 341-346.

Ward, J. (2006). *The StudentÕs Guide to Cognitive Neuroscience*. Sussex: Psychology Press.

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

Warrington, E.K., & Shallice, T. (1984). Category specific semantic impairments. *Brain*, 107, 829-854.

Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*. Breslau: Cohen & Weigart.

Werkera, J.F., & Tees, R.C. (2002). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, 25, 121-133.

Zatorre, R.J., Belin, P. & Penhune, V.B. (2002) Structure and function of auditory cortex: music and speech. *Trends in Cognitive Sciences*, 6, 37-46.