9

The Infrastructure of the Language-Ready Brain

Peter Hagoort and David Poeppel

Abstract

This chapter sketches in very general terms the cognitive architecture of both language comprehension and production, as well as the neurobiological infrastructure that makes the human brain ready for language. Focus is on spoken language, since that compares most directly to processing music. It is worth bearing in mind that humans can also interface with language as a cognitive system using sign and text (visual) as well as Braille (tactile); that is to say, the system can connect with input/output processes in any sensory modality. Language processing consists of a complex and nested set of subroutines to get from sound to meaning (in comprehension) or meaning to sound (in production), with remarkable speed and accuracy. The first section outlines a selection of the major constituent operations, from fractionating the input into manageable units to combining and unifying information in the construction of meaning. The next section addresses the neurobiological infrastructure hypothesized to form the basis for language processing. Principal insights are summarized by building on the notion of "brain networks" for speech–sound processing, syntactic processing, and the construction of meaning, bearing in mind that such a neat three-way subdivision overlooks important overlap and shared mechanisms in the neural architecture subserving language processing. Finally, in keeping with the spirit of the volume, some possible relations are highlighted between language and music that arise from the infrastructure developed here. Our characterization of language and its neurobiological foundations is necessarily selective and brief. Our aim is to identify for the reader critical questions that require an answer to have a plausible cognitive neuroscience of language processing.

The Cognitive Architecture of Language Comprehension and Production

The Comprehension of Spoken Language

When listening to speech, the first requirement is that the continuous speech input is perceptually segmented into discrete entities (features, segments,

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

syllables) that can be mapped onto, and will activate, abstract phonological representations that are stored in long-term memory. It is a common claim in state-of-the-art models of word recognition (the cohort model: Marslen-Wilson 1984; TRACE: McClelland and Elman 1986; the shortlist model: Norris 1994) that the incoming and unfolding acoustic input (e.g., the word-initial segment *ca*...) activates, in parallel, not only one but a whole set of lexical candidates (e.g., *captain, capture, captivate, capricious*…). This set of candidates is reduced, based on further incoming acoustic input and contextually based predictions, to the one that fits best (for a review, see Poeppel et al. 2008). This word recognition process happens extremely fast, and is completed within a few hundred milliseconds, whereby the exact duration is co-determined by the moment at which a particular word form deviates from all others in the mental lexicon of the listener (the so-called recognition point). Given the rate of typical speech (~4–6 syllables per second), we can deduce that word recognition is extremely fast and efficient, taking no more than 200–300 ms.

Importantly, achieving the mapping from acoustics to stored abstract representation is not the only subroutine in lexical processing. For example, words are not processed as unstructured, monolithic entities. Based on the morphophonological characteristics of a given word, a process of lexical decomposition takes place in which stems and affixes are separated. For spoken words, the trigger for decomposition can be something as simple as the inflectional rhyme pattern, which is a phonological pattern signaling the potential presence of an affix (Bozic et al. 2010). Interestingly, words seem to be decomposed by rule; that is, the decompositional, analytic processes are triggered for words with obvious parts (e.g., teacup = tea-cup; uninteresting = un-inter-est-ing) but also for semantically opaque words (e.g., bell-hop), and even nonwords with putative parts (e.g., blicket-s, blicket-ed). Decomposing lexical input appears to be a ubiquitous and mandatory perceptual strategy (e.g., Fiorentino and Poeppel 2007; Solomyak and Marantz 2010; and classic behavioral studies by Forster, Zwitserlood, Semenza, and others). Many relevant studies, especially with a view toward neurocognitive models, are reviewed by Marslen-Wilson (2007).

Recognizing word forms is an entrance point for the retrieval of syntactic (lemma) and semantic (conceptual) information. Here, too, the process is cascaded in nature. That is, based on partial phonological input, meanings of multiple lexical candidates are co-activated (Zwitserlood 1989). Multiple activation is less clear for lemma information that specifies the syntactic features (e.g., word class, grammatical gender) of a lexical entry. In most cases, the phrase structure context generates strong predictions about the syntactic slot (say, a noun or a verb) that will be filled by the current lexical item (Lau et al. 2006). To what degree lemma and concept retrieval are sequential or parallel in nature during online comprehension, is not clear. Results from electrophysiological recordings (event-related brain potential, ERP), however,

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

indicate that most of the retrieval and integration processes are completed within 500 ms (Kutas and Federmeier 2011; see also below).

Thus far, the processes discussed all relate to the retrieval of information from what is referred to in psycholinguistics as the mental lexicon. This is the information that in the course of language acquisition gets encoded and consolidated in neocortical memory structures, mainly located in the temporal lobes. However, language processing is (a) more than memory retrieval and (b) more than the simple concatenation of retrieved lexical items. The expressive power of human language (its generative capacity) derives from being able to combine elements from memory in endless, often novel ways. This process of deriving complex meaning from lexical building blocks (often called composition) will be referred to as *unification* (Hagoort 2005). As we will see later, (left) frontal cortex structures are implicated in unification.

In short, the cognitive architecture necessary to realize the expressive power of language is tripartite in nature, with levels of form (speech sounds, graphemes in text, or manual gestures in sign language), syntactic structure, and meaning as the core components of our language faculty (Chomsky 1965; Jackendoff 1999; Levelt 1999). These three levels are domain specific but, at the same time, they interact during incremental language processing. The principle of compositionality is often invoked to characterize the expressive power of language at the level of meaning. A strict account of compositionality states that the meaning of an expression is a function of the meanings of its parts and the way they are syntactically combined (Fodor and Lepore 2002; Heim and Kratzer 1998; Partee 1984). In this account, complex meanings are assembled bottom-up from the meanings of the lexical building blocks via the combinatorial machinery of syntax. This is sometimes referred to as simple composition (Jackendoff 1997). That some operations of this type are required is illustrated by the obvious fact that the same lexical items can be combined to yield different meanings: *dog bites man* is not the same as *man bites dog*. Syntax matters. It matters, however, not for its own sake but in the interest of mapping grammatical roles (subject, object) onto thematic roles (agent, patient) in comprehension, and in the reverse order in production. The thematic roles will fill the slots in the situation model (specifying states and events) representing the intended message.

That this account is not sufficient can be seen in adjective–noun expressions such as *flat tire*, *flat beer*, *flat note*, etc. (Keenan 1979). In all these cases, the meaning of "flat" is quite different and strongly context dependent. Thus, structural information alone will need to be supplemented. On its own, it does not suffice for constructing complex meaning on the basis of lexical-semantic building blocks. Moreover, ERP (and behavioral) studies have found that nonlinguistic information which accompanies the speech signal (such as information about the visual environment, about the speaker, or about co-speech gestures; Van Berkum et al. 2008; Willems et al. 2007; Willems et al. 2008) are unified in parallel with linguistic sources of information. Linguistic and

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

nonlinguistic information conspire to determine the interpretation of an utterance on the fly. This all happens extremely fast, usually in less than half a second. For this and other reasons, simple (or strict) composition seems not to hold across all possible expressions in the language (see Baggio and Hagoort 2011).

We have made a distinction between memory retrieval and unification operations. Here we sketch in more detail the nature of unification in interaction with memory retrieval. Classically, psycholinguistic studies of unification have focused on syntactic analysis. However, as we saw above, unification operations take place not only at the syntactic processing level. Combinatoriality is a hallmark of language across representational domains (cf. Jackendoff 2002). Thus, at the semantic and phonological levels, too, lexical elements are argued to be combined and integrated into larger structures (cf. Hagoort 2005). Nevertheless, models of unification are most explicit for syntactic processing. For this level of analysis, we can illustrate the distinction between memory retrieval and unification most clearly, According to the *memory, unification, and control* (MUC) model (Hagoort 2005), each word form in the mental lexicon is associated with a structural frame (Vosse and Kempen 2000). This structural frame consists of a three-tiered unordered tree, specifying the possible structural environment of the particular lexical item (see Figure 9.1).

The top layer of the frame consists of a single phrasal node (e.g., noun phrase, NP). This so-called root node is connected to one or more functional nodes (e.g., subject, S; head, hd; direct object, dobj) in the second layer of the frame. The third layer again contains phrasal nodes to which lexical items or other frames can be attached.

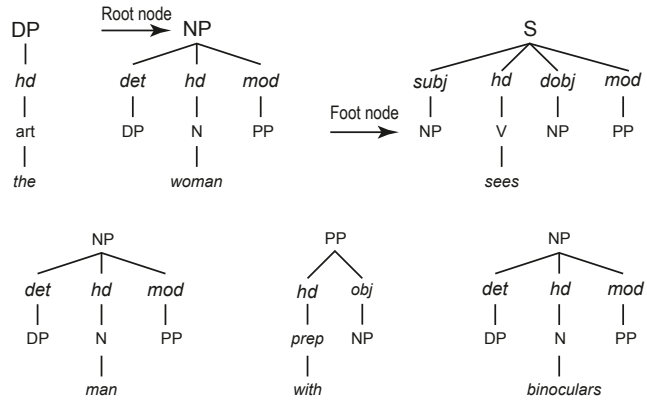

Figure 9.1 Syntactic frames in memory. Frames such as these are retrieved on the basis of incoming word form information (*the, woman,* etc). DP: determiner phrase; NP: noun phrase; S: sentence; PP: prepositional phrase; art: article; hd: head; det: determiner; mod: modifier; subj: subject; dobj: direct object. The head of a phrase determines the syntactic type of the frame (e.g., noun for a noun phrase, preposition for a prepositional phrase)

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

This parsing account is "lexicalist" in the sense that all syntactic nodes—S, NP, VP (verb phrase), N, V—are retrieved from the mental lexicon. In other words, chunks of syntactic structure are stored in memory. There are no syntactic rules that introduce additional nodes, such as in classical rewrite rules in linguistics (S → NP VP). In the online comprehension process, structural frames associated with the individual word forms incrementally enter the unification workspace. In this workspace, constituent structures spanning the whole utterance are formed by a unification operation (see Figure 9.2). This operation consists of linking up lexical frames with identical root and foot nodes, and checking agreement features (number, gender, person, etc.). Although the lexical-syntactic frames might differ between languages, as well as the ordering of the trees, what is claimed to be universal is the combination of lexically specified syntactic templates and unification procedures. Moreover, across language the same distribution of labor is predicted between brain areas involved in memory and brain areas that are crucial for unification.

The resulting unification links between lexical frames are formed dynamically, which implies that the strength of the unification links varies over time until a state of equilibrium is reached. Due to the inherent ambiguity in natural language, alternative unification candidates will usually be available at any point in the parsing process. That is, a particular root node (e.g., prepositional phrase, PP) often finds more than one matching foot node (i.e., PP) (see Figure 9.2) with which it can form a unification link (for examples, see Hagoort 2003).

Ultimately, at least for sentences which do not tax the processing resources very strongly, one phrasal configuration results. This requires that among the alternative binding candidates, only one remains active. The required state of equilibrium is reached through a process of lateral inhibition between two or

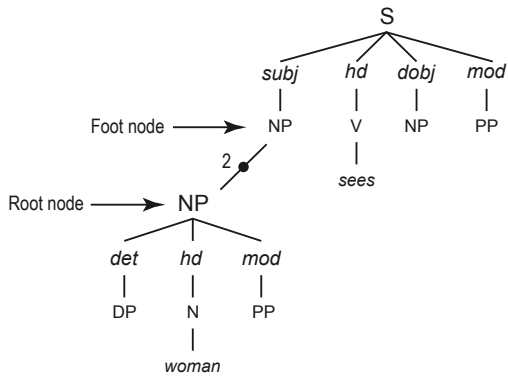

Figure 9.2 The unification operation of two lexically specified syntactic frames. Unification takes place by linking the root node NP to an available foot node of the same category. The number 2 indicates that this is the second link that is formed during online processing of the sentence, *The woman sees the man with the binoculars*.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

more alternative unification links. In general, due to gradual decay of activation, more recent foot nodes will have a higher level of activation than the ones that entered the unification space earlier. In addition, strength levels of the unification links can vary as a function of plausibility (semantic) effects. For instance, if instrumental modifiers under S-nodes have a slightly higher default activation than instrumental modifiers under an NP-node, lateral inhibition can result in overriding a recency effect.

The picture that we sketched above is based on the assumption that we always create a fully unified structure. This is, however, unlikely. In our actual online processing of life in a noisy world, the comprehension system will often work with just bits and pieces (e.g., syntactic frames) that are not all unified into one fully unified phrasal configuration. Given both extralinguistic and language-internal contextual prediction and redundancy, in the majority of cases this is still good enough to derive the intended message (see below).

The *unification model*, as formalized in Vosse and Kempen (2000), has nevertheless a certain psychological plausibility. It accounts for sentence complexity effects known from behavioral measures, such as reading times. In general, sentences are harder to analyze syntactically when more potential unification links of similar strength enter into competition with each other. Sentences are easy when the number of U-links is small and of unequal strength. In addition, the model accounts for a number of other experimental findings in psycholinguistic research on sentence processing, including syntactic ambiguity (attachment preferences; frequency differences between attachment alternatives), and lexical ambiguity effects. Moreover, it accounts for breakdown patterns in agrammatic sentence analysis (for details, see Vosse and Kempen 2000).

So far we have specified the memory and retrieval operations that are triggered by the orthographic or acoustic input. Similar considerations apply to sign language. In our specification of the processing steps involved, we have implicitly assumed that ultimately decoding the meaning is what language comprehension is about. However, while this might be a necessary aspect, it cannot be the whole story. Communication goes further than the exchange of explicit propositions. In essence, it is a way to either change the mind of the listener, or to commit the addressee to the execution of certain actions, such as closing the window in reply to the statement *It is cold in here*. In other words, a theory of speech acts is required to understand how we get from coded meaning to inferred speaker meaning (cf. Levinson, this volume; Grice 1989).

Another assumption that we made, but which might be incorrect, relates to how much of the input the listener/reader analyzes. This is what we alluded to briefly in the context of unification. In classical models of sentence comprehension—of either the syntactic structure-driven variety (Frazier 1987) or in a constraint-based framework (Tanenhaus et al. 1995)—the implicit assumption is usually that a full phrasal configuration results and a

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

complete interpretation of the input string is achieved. However, oftentimes the listener interprets the input on the basis of bits and pieces that are only partially analyzed. As a consequence, the listener might overhear semantic information (cf. the Moses illusion; Erickson and Mattson 1981) or syntactic information (cf. the Chomsky illusion; Wang et al. 2012). In the question *How many animals of each kind did Moses take on the ark?*, people often answer "two," without noticing that it was Noah who was the guy with an ark, and not Moses. Likewise, we found that syntactic violations might go unnoticed if they are in a sentence constituent that provides no new information (Wang et al. 2012). Ferreira et al. (2002) introduced the phrase *good-enough processing* to refer to the listeners' and readers' interpretation strategies. In a good-enough processing context, linguistic devices that highlight the most relevant parts of the input might help the listener/reader in allocating processing resources optimally. This aspect of linguistic meaning is known as *information structure* (Büring 2007; Halliday 1967). The information structure of an utterance essentially focuses the listener's attention on the crucial (new) information in it. In languages such as English and Dutch, prosody plays a crucial role in marking information structure. For instance, in question–answer pairs, the new or relevant information in the answer will typically be pitch accented. After a question like *What did Mary buy at the market?* the answer might be *Mary bought VEGETABLES* (accented word in capitals). In this case, the word "vegetables" is the focus constituent, which corresponds to the information provided for the Wh-element in the question. There is no linguistic universal for signaling information structure. The way information structure is expressed varies within and across languages. In some languages it may impose syntactic locations for the focus constituent; in others focus-marking particles are used, or prosodic features like phrasing and accentuation (Kotschi 2006; Miller 2006).

In summary, language comprehension requires an analysis of the input that allows the retrieval of relevant information from memory (the mental lexicon). The lexical building blocks are unified into larger structures decoding the propositional content. Further inferential steps are required to derive the intended message of the speaker from the coded meaning. Based on the listener's comprehension goals, the input is analyzed to a lesser or greater degree. Linguistic marking of information structure co-determines the depth of processing of the linguistic input. In addition, nonlinguistic input (e.g., co-speech gestures, visual context) is immediately integrated into the situation model that results from processing language in context.

Producing Language

While speech comprehension can be described as the mapping from sound (or sign) to meaning, in speaking we travel the processing space in the reverse order. In speaking, a preverbal message is transformed in a series of steps into

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

a linearized sequence of speech sounds (for details, see Levelt 1989, 1999). This again requires the retrieval of building blocks from memory and their unification at multiple levels. Most research on speaking has focused on single word production, as in picture naming. The whole cascade of processes, from the stage of conceptual preparation to the final articulation, happens in about 600 ms (Indefrey and Levelt 2004). Since we perform this process in an incremental fashion, we can easily utter 2–4 words per second. Moreover, this is done with amazing efficiency; on average, a speech error occurs only once in a thousand words (Bock 2011; Deese 1984). The whole cascade of processes starts with the preverbal message, which triggers the selection of the required lexical concepts (i.e., the concepts for which a word form is available in the mental lexicon). The activation of a lexical concept leads to the retrieval of multiple lemmas and a selection of the target lemma, which gets phonologically encoded. At the stage of lemma selection, morphological unification of, for instance, stem and affix takes place. Recent intracranial recordings in humans indicate that certain parts of Broca's region are involved in this unification process (Sahin et al. 2009). Once the phonological word forms are retrieved, they will result in the retrieval and unification of the syllables that compose a phonological word in its current speech context.

Although speech comprehension and speaking recruit many of the same brain areas during sentence-level semantic processes, syntactic operations, and lexical retrieval (Menenti et al. 2011), there are still important differences. The most important difference is that although speakers pause, repair, etc., they nevertheless cannot bypass syntactic and phonological encoding of the utterance that they intend to produce. What is good enough for the listener is often not good enough for the speaker. Here, the analogy between perceiving and producing music seems obvious. It may well be that the interconnectedness of the cognitive and neural architectures for language comprehension and production enables the production system to participate in generating internal predictions while in the business of comprehending linguistic input. This prediction-is-production account, however, may not be as easy in relation to the perception of music, at least for instrumental music. With few exceptions, all of humankind are expert speakers. However, for music, there seems to be a stronger asymmetry between perception and production. This, then, results in two questions: Does prediction play an equally strong role in language comprehension and the perception of music? If so, what might generate the predictions in music perception?

The Neurobiological Infrastructure

Classically, and based primarily on evidence from deficits in aphasic patients, the perisylvian cortex in the left hemisphere has been seen as the crucial network for supporting the processing of language. The critical components

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

were assumed to be Broca's area in the left inferior frontal cortex (LIFC) and Wernicke's area in the left superior temporal cortex, with these areas mutually connected by the arcuate fasciculus. These areas, and their roles in language comprehension and production, are often still described as the core language nodes in handbooks on brain function (see Figure 9.3).

However, later patient studies, and especially recent neuroimaging studies in healthy subjects, have revealed that (a) the distribution of labor between Broca's and Wernicke's areas is different than proposed in the classical model, and (b) a much more extended network of areas is involved, not only in the left hemisphere, but also involving homologous areas in the right hemisphere. One alternative proposal is the MUC model proposed by Hagoort (2005). In this model, the distribution of labor is as follows (see Figure 9.4): Areas in the temporal cortex (in yellow) subserve the knowledge representations that have been laid down in memory during acquisition. These areas store information about word form, word meanings, and the syntactic templates that we discussed above. Dependent on information type, different parts of temporal cortex are involved. Frontal cortex areas (Broca's area and adjacent cortex, in blue) are crucial for the unification operations. These operations generate larger structures from the building blocks that are retrieved from memory. In addition, executive control needs to be exerted, such that the correct target

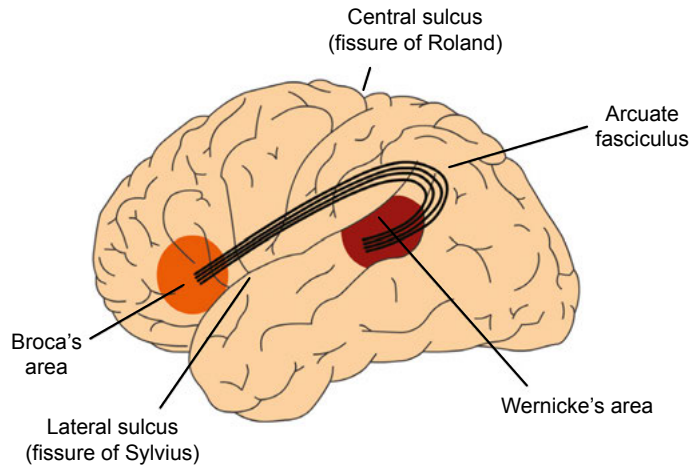

Figure 9.3 The classical Wernicke–Lichtheim–Geschwind model of the neurobiology of language. In this model Broca's area is crucial for language production, Wernicke's area subserves language comprehension, and the necessary information exchange between these areas (such as in reading aloud) is done via the arcuate fasciculus, a major fiber bundle connecting the language areas in temporal cortex (Wernicke's area) and frontal cortex (Broca's area). The language areas border one of the major fissures in the brain, the so-called Sylvian fissure. Collectively, this part of the brain is often referred to as perisylvian cortex.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

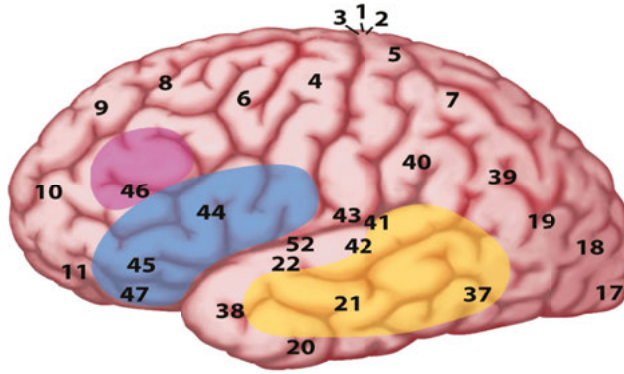

Figure 9.4 The MUC model of language. The figure displays a lateral view of the left hemisphere. The numbers indicate Brodmann areas. These are areas with differences in the cytoarchitectonics (i.e., composition of cell types). The memory areas are in the temporal cortex (in yellow). Unification requires the contribution of Broca's area (Brodmann areas 44 and 45) and adjacent cortex (Brodmann areas 47 and 6) in the frontal lobe. Control operations recruit another part of the frontal lobe (in pink) and the anterior cingulate cortex (not shown in the figure).

language is selected, turn-taking in conversation is orchestrated, etc. Control areas involve dorsolateral prefrontal cortex (in pink) and a midline structure known as the anterior cingulate cortex (not shown in Figure 9.4).

In the following sections we discuss in more detail the brain networks which support the different types of information that are crucial for language. We briefly describe the neurobiological infrastructure underlying the tripartite architecture of the human language system. For the three core types of information (phonological, syntactic, and semantic), we make the same general distinction between retrieval operations and unification: Retrieval refers to accessing language-specific information in memory. Unification is the (de)composition of larger structures from the building blocks that are retrieved from memory. As we will see below, a similar distinction has been proposed for music, with a striking overlap in the recruitment of the neural unification network for language and music (Patel 2003 and this volume).

The Speech and Phonological Processing Network

As we noted at the outset, speech perception is not an unstructured, monolithic cognitive function. Mapping from sounds to words involves multiple steps, including operations that depend on what one is expected to do as a listener: remain silent (passive listening), repeat the input, write it down, etc. The different tasks will play a critical role in the perception process. Accordingly, it is now well established that there is no single brain area that is responsible for speech perception and the activation/recruitment of phonological knowledge.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

Rather, several brain regions in different parts of the cerebral cortex interact in systematic ways in speech perception. The overall network, which also includes subcortical contributions (see recent work by Kotz and Schwartze 2010), has been established by detailed consideration of brain injury and functional imaging data (for reviews and perspectives on this, see Binder 2000; Hickok and Poeppel 2000, 2004, 2007; Poeppel et al. 2008; Scott and Johnsrude 2003). Figure 9.5, from Hickok and Poeppel (2007), summarizes one such perspective, emphasizing concurrent processing pathways.

Areas in the temporal lobe, parietal areas, and several frontal regions conspire to form the network for speech recognition. The functional anatomy underlying speech–sound processing is comprised of a distributed cortical system that encompasses regions along at least two processing streams. A ventral, temporal lobe pathway (see Figure 9.5b) primarily mediates the mapping from sound input to meaning/words (lower pathway in Figure 9.5a). A dorsal path incorporating parietal and frontal lobes enables the sensorimotor transformations that underlie

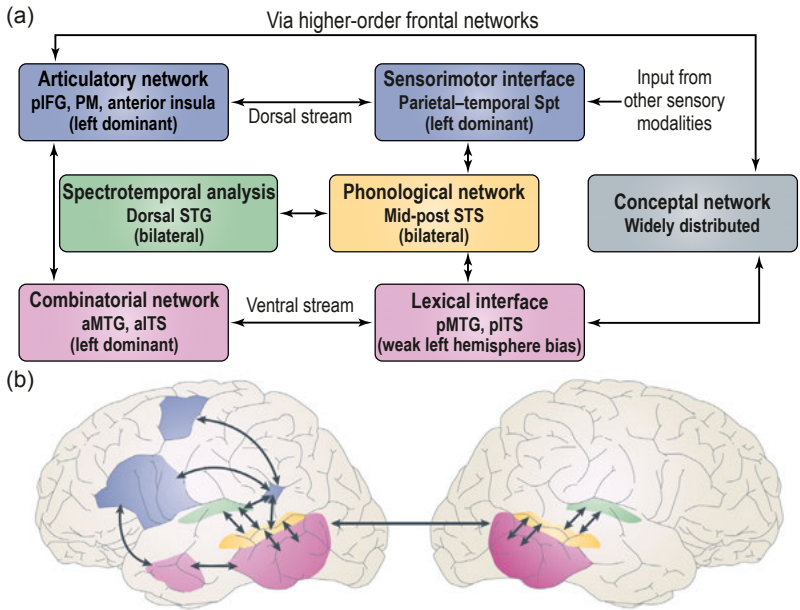

Figure 9.5 A model of the speech and phonological processing network. The earliest stages of cortical speech processing involve some form of spectrotemporal analysis, which is carried out in auditory cortices bilaterally in the supratemporal plane. Phonological-level processing and representation involves the middle to posterior portions of the superior temporal sulcus (STS) bilaterally, although there might be a left-hemisphere bias at this level of processing. A dorsal pathway (blue) maps sensory or phonological representations onto articulatory motor representations. A ventral pathway (pink) provides the interface with memory representations of lexical syntax and lexical concepts (reprinted with permission from Hickok and Poeppel 2007).

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

mapping to output representations (upper pathway in Figure 9.5a). This anatomic fractionation suggests that hypothesized subroutines and representations of speech processing have their own neural realization, as indicated in the boxes, and supports models which posit a componential architecture (e.g., this dual pathway model). This distributed functional anatomy for speech recognition contrasts with other systems. For example, in the study of face recognition, one brain region plays a disproportionately large role (the fusiform face area). However, the functional anatomic models that have been developed for speech recognition and phonological processing are much more extended and bear a resemblance to the organization of the visual system. In the parallel pathways in the visual system, we contrast a where/how (dorsal) and a what (ventral) system (Kravitz et al. 2011).

One way to carve up the issue—admittedly superficial, but mnemonically useful—is purely by anatomy: temporal lobe–memory; parietal lobe–analysis/coordinate transformation; frontal lobe–synthesis/unification (Ben Shalom and Poeppel 2008). The areas in the temporal lobe (in addition to sensory/perceptual analysis in the superior temporal lobe) have a principal role in storage and retrieval of speech sounds and words. These areas underlie the required memory functions. One region in the temporal lobe of special interest in the mapping from sound form to lexical representation is the superior temporal sulcus (STS); it receives inputs from many areas, including core auditory fields, visual areas, etc., and it sits adjacent to middle temporal gyrus (MTG), the putative site of lexical representations proper (Hickok and Poeppel 2004; Indefrey and Levelt 2004; Lau et al. 2006; Snijders et al. 2009). The areas in the parietal cortex (SPT, SMG, angular gyrus, intraparietal sulcus) are implicated in analytic functions (e.g., sublexical phonological decisions; sensorimotor transformations). The areas in frontal cortex (various areas in the inferior frontal cortex and dorsomedial frontal cortex) play an obvious role in setting up motor output programming, but, more critically, underlie unification operations.

The Syntactic Network

In comparison with phonological and semantic processing, which have compelling bilateral contributions (in contrast to the classical left-hemisphere-only model), syntactic processing seems strongly lateralized to the left hemisphere perisylvian regions. Indirect support for a distinction between a memory component (i.e., the mental lexicon) and a unification component in syntactic processing comes from neuroimaging studies on syntactic processing. In a meta-analysis of 28 neuroimaging studies, Indefrey (2004) found two areas that were critical for syntactic processing, independent of the input modality (visual in reading, auditory in speech). These two supramodal areas for syntactic processing were the left posterior STG/MTG and the LIFC. The left posterior temporal cortex is known to be involved in lexical

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

processing (Hickok and Poeppel 2004, 2007; Indefrey and Cutler 2004; Lau et al. 2006). In connection to the *unification model*, this part of the brain might be important for the retrieval of the syntactic frames that are stored in the lexicon. The *unification space*, where individual frames are connected into a phrasal configuration for the whole utterance, might recruit the contribution of Broca's area (LIFC).

Direct empirical support for this distribution of labor between LIFC (Broca's area) and temporal cortex was recently found in a study of Snijders et al. (2009). These authors performed an fMRI study in which participants read sentences and word sequences containing word–category (noun–verb) ambiguous words at critical position (e.g., "watch"). Regions contributing to the syntactic unification process should show enhanced activation for sentences compared with words, and only within sentences display a larger signal for ambiguous than unambiguous conditions. The posterior LIFC showed exactly this predicted pattern, confirming the hypothesis that LIFC contributes to syntactic unification. The left posterior MTG was activated more for ambiguous than unambiguous conditions, as predicted for regions subserving the retrieval of lexical-syntactic information from memory. It thus seems that the LIFC is crucial for syntactic processing in conjunction with the left posterior MTG, a finding supported by patient studies with lesions in these very same areas (Caplan and Waters 1996; Rodd et al. 2010; Tyler et al. 2011).

In the domain of music perception, a similar model has been proposed by Patel (2003). Although in the past, perspectives on language and music often stressed the differences, Patel has introduced and strongly promotes an alternative view: that at many levels, the similarities between music and language are more striking than the differences. Clearly, the differences are undeniable. For instance, there are pitch intervals in music that we do not have in language; on the other hand, nouns and verbs are part of the linguistic system without a concomitant in music. These examples point to differences in the representational structures that are domain specific and laid down in memory during acquisition. However, the processing mechanisms (algorithms) and the neurobiological infrastructure to retrieve and combine domain-specific representations might be shared to a large extent. This idea has been made explicit in Patel's *shared syntactic integration resource hypothesis* (SSIRH in short; see Patel, this volume). According to this hypothesis, linguistic and musical syntax have mechanisms of sequencing in common, which are instantiated in overlapping frontal brain areas that operate on different domain-specific syntactic representations in posterior brain regions. Patel's account predicts that lesions affecting the unification network in patients with Broca's aphasia should also impair their unification capacity for music. In fact, this is exactly what a collaborative research project between Patel's and Hagoort's research groups has found (Patel et al. 2008a).

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

The Semantic Network

In recent years, there has been growing interest in investigating the cognitive neuroscience of semantic processing (for a review of a number of different approaches, see Hinzen and Poeppel 2011). A series of fMRI studies has aimed at identifying the semantic processing network. These studies either compared sentences containing semantic/pragmatic anomalies with their correct counterparts (e.g., Friederici et al. 2003; Hagoort et al. 2004; Kiehl et al. 2002; Ruschemeyer et al. 2006) or sentences with and without semantic ambiguities (Davis et al. 2007; Hoenig and Scheef 2005; Rodd et al. 2005). The most consistent finding across all of these studies is the activation of the LIFC, in particular BA 47 and BA 45. In addition, the left superior and middle temporal cortices are often found to be activated, as well as left inferior parietal cortex. For instance, Rodd and colleagues had subjects listen to English sentences such as *There were dates and pears in the fruit bowl* and compared the fMRI response of these sentences to the fMRI response of sentences such as *There was beer and cider on the kitchen shelf*. The crucial difference between these sentences is that the former contains two homophones, i.e., "dates" and "pears," which, when presented auditorily, have more than one meaning. This is not the case for the words in the second sentence. The sentences with the lexical ambiguities led to increased activations in LIFC and in the left posterior middle/inferior temporal gyrus. In this experiment all materials were well-formed English sentences in which the ambiguity usually goes unnoticed. Nevertheless, very similar results were obtained in experiments that used semantic anomalies. Areas involved in semantic unification were found to be sensitive to the increase in semantic unification load due to the ambiguous words.

Semantic unification could be seen as filling the slots in an abstract event schema, where in the case of multiple word meanings for a given lexical item competition and selection increase in relation to filling a particular slot in the event scheme. As with syntactic unification, the availability of multiple candidates for a slot will increase the unification load. In the case of the lexical ambiguities there is no syntactic competition, since both readings activate the same syntactic template (in this case the NP-template). Increased processing is hence due to integration of meaning instead of syntax.

In short, the semantic processing network seems to include at least LIFC, left superior/middle temporal cortex, and the (left) inferior parietal cortex. To some degree, the right hemisphere homologs of these areas are also found to be activated. Below we will discuss the possible contributions of these regions to semantic processing.

An indication for the respective functional roles of the left frontal and temporal cortices in semantic unification comes from a few studies investigating semantic unification of multimodal information with language. Using fMRI, Willems and colleagues assessed the neural integration of semantic information from spoken words and from co-speech gestures into a preceding sentence

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

context (Willems et al. 2007). Spoken sentences were presented in which a critical word was accompanied by a co-speech gesture. Either the word or the gesture could be semantically incongruous with respect to the previous sentence context. Both an incongruous word as well as an incongruous gesture led to increased activation in LIFC as compared to congruous words and gestures (for a similar finding with pictures of objects, see Willems et al. 2008). Interestingly, the activation of the left posterior STS was increased by an incongruous spoken word, but not by an incongruous hand gesture. The latter resulted in a specific increase in dorsal premotor cortex (Willems et al. 2007). This suggests that activation increases in left posterior temporal cortex are triggered most strongly by processes involving the retrieval of lexical-semantic information. LIFC, on the other hand, is a key node in the semantic unification network, unifying semantic information from different modalities. From these findings it seems that semantic unification is realized in a dynamic interplay between LIFC as a multimodal unification site, on the one hand, and modality-specific areas on the other.

Although LIFC (including Broca's area) has traditionally been construed as a language area, a wealth of recent neuroimaging data suggests that its role extends beyond the language domain. Several authors have thus argued that LIFC function is best characterized as "controlled retrieval" or "(semantic) selection" (Badre et al. 2005; Moss et al. 2005; Thompson-Schill et al. 2005; Thompson-Schill et al. 1997; Wagner et al. 2001). How does the selection account of LIFC function relate to the unification account? As discussed elsewhere, unification often implies selection (Hagoort 2005). For instance, in the study by Rodd and colleagues described above, increased activation in LIFC is most likely due to increased selection demands in reaction to sentences with ambiguous words. Selection is often, but not always, a prerequisite for unification. Unification with or without selection is a core feature of language processing. During natural language comprehension, information has to be kept in working memory for a certain period of time, and incoming information is integrated and combined with previous information. The combinatorial nature of language necessitates that a representation is constructed online, without the availability of an existing representation of the utterance in long-term memory. In addition, some information sources that are integrated with language do not have a stable representation in long-term memory such that they can be selected. For instance, there is no stable representation of the meaning of co-speech gestures, which are highly ambiguous outside of a language context. Still, in all these cases increased activation is observed in LIFC, such as when the integration load of information from co-speech gestures is high (Willems et al. 2007). Therefore, unification is a more general account of LIFC function. It implies selection, but covers additional integration processes as well.

Importantly, semantic processing is more than the concatenation of lexical meanings. Over and above the retrieval of individual word meanings, sentence and discourse processing requires combinatorial operations that result in

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

a coherent interpretation of multi-word utterances. These operations do not adhere to a simple principle of compositionality alone. World knowledge, information about the speaker, co-occurring visual input, and discourse information all trigger similar electrophysiological responses as sentence-internal semantic information. A network of brain areas, including the LIFC, the left superior/middle/inferior temporal cortex, the left inferior parietal cortex and, to a lesser extent, their right hemisphere homologs, are recruited to perform semantic unification. The general finding is that semantic unification operations are under top-down control of the left and, in the case of discourse, also the right inferior frontal cortex. This contribution modulates activations of lexical information in memory as represented by the left superior and middle temporal cortex, with presumably additional support for unification operations in left inferior parietal areas (e.g., angular gyrus).

The Network Topology of the Language-Ready Brain

We have seen that the language network in the brain is much more extended than was thought for a long time and includes areas in the left hemisphere as well as right hemisphere. However, the evidence of additional activations in the right hemisphere and areas other than Broca and Wernicke does not take away the strong bias in favor of left perisylvian cortex. In a recent meta-analysis based on 128 neuroimaging papers, Vigneau et al. (2010) compared left and right hemisphere activations that were observed in relation to language processing. On the whole, for phonological processing, lexical-semantic processing, and sentence or text processing, the activation peaks in the right hemisphere comprised less than one-third of the activation peaks in the left hemisphere. Moreover, in the large majority of cases, right hemisphere activations were in homotopic areas, suggesting strong interhemispheric influence. It is therefore justified to think that for the large majority of the population (with the exception of some portion of left-handers, cases of left hemispherectomy, etc.), the language readiness of the human brain resides to a large extent in the organization of the left perisylvian cortex. One emerging generalization is that the network of cortical regions subserving output processing (production) is very strongly (left) lateralized; in contrast, the computational subroutines underlying comprehension appear to recruit both hemispheres rather extensively, even though there also exists compelling lateralization, especially for syntax (Menenti et al. 2011).

Moreover, the network organization of the left perisylvian cortex has been found to show characteristics that distinguishes it from the right perisylvian cortex and from homolog areas in other primates. A recent technique for tracing fiber bundles in the living brain is diffusion tensor imaging (DTI). Using DTI, Rilling et al. (2008) tracked the arcuate fasciculus in humans, chimpanzees, and macaques and found a prominent temporal lobe projection of the arcuate

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

fasciculus in humans that is much smaller or absent in nonhuman primates (see Figure 9.6). Moreover, connectivity with the MTG was more widespread and of higher probability in the left than in the right hemisphere. This human specialization may be relevant for the evolution of language. Catani et al. (2007) found that the human arcuate fasciculus is strongly lateralized to the left, with quite some variation on the right. On the right, some people lack an arcuate fasciculus, in others it is smaller in size, and in a minority of the population this fiber bundle is of equal size in both hemispheres. The presence of the arcuate fasciculus in the right hemisphere correlated with a better verbal memory. This pattern of lateralization was confirmed in a study on 183 healthy right-handed volunteers aged 5–30 years (Lebel and Beaulieu 2009). In this study the lateralization pattern did not differ with age or gender. The arcuate fasciculus lateralization is present at five years of age and remains constant

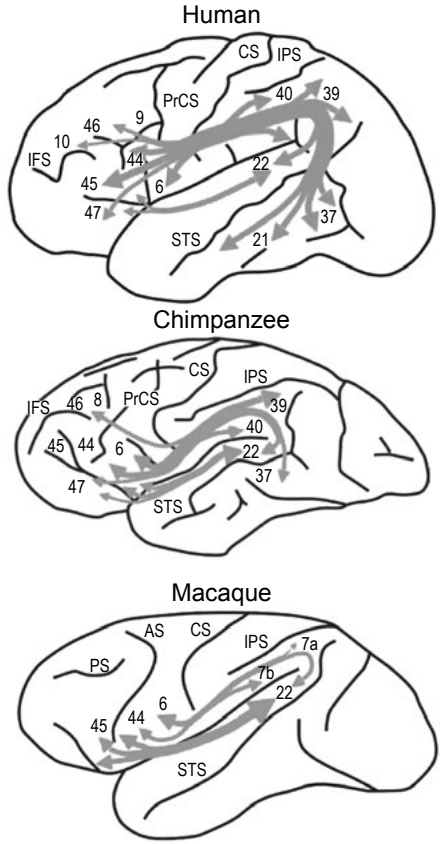

Figure 9.6 The arcuate fasciculus in human, chimpanzee, and macaque in a schematic lateral view of the left hemisphere. Reprinted from Rilling et al. (2008) with permission from Macmillan Publishers Ltd.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

throughout adolescence into adulthood. However, another recent study comparing seven-year-olds with adults (Brauer et al. 2011b) shows that the arcuate fasciculus is still relatively immature in the children.

In addition to the arcuate fasciculus, which can be viewed as part of a dorsal processing stream, other fiber bundles are also important in connecting frontal with temporoparietal language areas (see Figure 9.7). These include the superior longitudinal fasciculus (adjacent to the arcuate fasciculus) and the extreme capsule fasciculus as well as the uncinate fasciculus, connecting Broca's area with superior and middle temporal cortex along a ventral path (Anwander et al. 2007; Friederici 2009a; Kelly et al. 2010).

DTI is not the only way to trace brain connectivity. It has been found that imaging the brain during rest reveals low-frequency (<0.1 Hz) fluctuations in the fMRI signal. It turns out that these fluctuations are correlated across areas that are functionally related (Biswal et al. 1995; Biswal and Kannurpatti 2009). This so-called resting state fMRI can thus be used as an index of functional connectivity. Although both DTI and resting state fMRI measure connectivity, in the case of DTI the connectivity can often be related to anatomically identifiable fiber bundles. Resting state connectivity measures the functional correlations between areas without providing a correlate in terms of an anatomical tract. Using the resting state method, Xiang et al. (2010) found a clear topographical functional connectivity pattern in the left inferior frontal, parietal, and temporal areas. In the left but not the right perisylvian cortex, functional connectivity patterns obeyed the tripartite nature

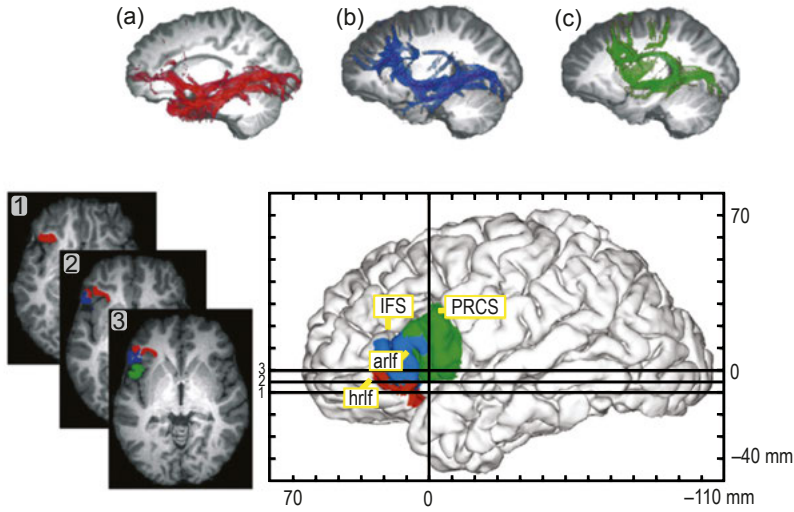

Figure 9.7 Connectivity patterns between parts of frontal cortex (in red, blue, and green) with parietal and temporal areas. Colored areas in left frontal cortex are connected via fiber bundles with the same color (a, b, c). Reprinted from Friederici (2009a) with permission from Elsevier.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

of language processing (phonology, syntax, and semantics). These results support the assumption of the functional division for phonology, syntax, and semantics of the LIFC, including Broca's area, as proposed by the MUC model (Hagoort 2005), and revealed a topographical functional organization in the left perisylvian language network, in which areas are most strongly connected according to information type (i.e., phonological, syntactic, and semantic).

In summary, despite increasing evidence of right hemisphere involvement in language processing, it still seems clear that the left perisylvian cortex has certain network features that stand out in comparison to other species, making it especially suited for supporting the tripartite architecture of human language.

Neurophysiology and Timing

Although we have thus far emphasized functional neuroanatomy and the insights from imaging, it is worth bearing in mind what electrophysiological data add to the functional interpretations we must entertain. As discussed at the outset, one of the most remarkable characteristics of speaking and listening is the speed at which it occurs. Speakers produce easily between two and five words per second; information that has to be decoded by the listener within roughly the same time frame. Considering that the acoustic duration of many words is in the order of a few hundred milliseconds, the immediacy of the electrophysiological language-related effects is remarkable. For instance, the early left anterior negativity (ELAN), a syntax-related effect (Friederici et al. 2003), has an onset on the order 100–150 ms after the acoustic word onset. The onset of the N400 is approximately at 250 ms, and another language relevant ERP, the so-called P600, usually starts at about 500 ms. Thus the majority of these effects happen well before the end of a spoken word. Classifying visual input (e.g., a picture) as depicting an animate or inanimate entity takes the brain approximately 150 ms (Thorpe et al. 1996). Roughly the same amount of time is needed to classify orthographic input as a letter (Grainger et al. 2008). If we take this as our reference time, the early appearance of an ELAN response to a spoken word is remarkable, to say the least. In physiological terms, it might be just too fast for long-range recurrent feedback to have its effect on parts of primary and secondary auditory cortex involved in first-pass acoustic and phonological analysis. Recent modeling work suggests that early ERP effects are best explained by a model with feed-forward connections only. Backward connections become essential only after 220 ms (Garrido et al. 2007). The effects of backward connections are, therefore, not manifest in the latency range of at least the ELAN, since not enough time has passed for return activity from higher levels. However, in the case of speech, the N400 follows the word recognition points closely in time. This suggests that what is happening in online language comprehension is presumably, for a substantial part, based on predictive processing. Under most circumstances, there is simply not enough time for top-down feedback to exert control over a

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

preceding bottom-up analysis. Very likely, lexical, semantic, and syntactic cues conspire to predict very detailed characteristics of the next anticipated word, including its syntactic and semantic makeup. A mismatch between contextual prediction and the output of bottom-up analysis results in an immediate brain response recruiting additional processing resources for the sake of salvaging the online interpretation process. Recent ERP studies have provided evidence that context can indeed result in predictions about a next word's syntactic features (i.e., gender; Van Berkum et al. 2005) and word form (DeLong et al. 2005). Lau et al. (2006) provided evidence that the ELAN elicited by a word category violation was modulated by the strength of the expectation for a particular word category in the relevant syntactic slot. In summary, we conclude that predictive coding is likely a central feature of the neurocognitive infrastructure.

Neural Rhythms and the Structure of Speech

A final issue relates to the convergence of intrinsic aspects of brain function and temporal characteristics of the speech signal. It is known that the brain generates intrinsic oscillatory rhythms which can be characterized by their frequency bands (for an extended discussion of the neural underpinnings, see Buzsáki 2006). For instance, theta oscillations are defined as activity between ~4 and 8 Hz, the alpha rhythm has its center peak at about 10 Hz (~9–12 Hz), and beta oscillations are found at around 20 Hz. Finally, gamma oscillations are characterized by frequencies above 40 Hz (see also Arbib, Verschure, and Seifert, this volume.) A recent, and admittedly still speculative, hypothesis suggests the intriguing possibility that some of these neuronal oscillations have temporal properties that make them ideally suited to be the carrier waves for processing aspects of language that are characterized by the different timescales at which they occur (e.g., Giraud et al. 2007; Luo and Poeppel 2007; Schroeder et al. 2008; Giraud and Poeppel 2012).

Naturalistic, connected speech is aperiodic, but nevertheless quasi-rhythmic as an acoustic signal. This temporal regularity in speech occurs at multiple timescales; each of these scales is associated with different types of perceptual information in the signal. Very rapidly modulated information, say 30–40 Hz or above (low gamma band), is associated with the spectrotemporal fine structure of a signal and is critical for establishing the order of rapid events. Modulation at the rate of 4–8 Hz (the so-called theta band) is associated with envelope fluctuations, discussed below. Modulations at slow rates, say 1–3 Hz, typically signal prosodic aspects of utterances, including intonation contour and phrasal attributes. We briefly elaborate on one of these scales: the intermediate scale.

There exists one pronounced temporal regularity in the speech signal at relatively low modulation frequencies. These modulations of signal energy (in reality, spread out across a filter bank) are well below 20 Hz, typically peaking roughly at a rate of 4–6 Hz. From the perspective of what the auditory cortex

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

receives as input, namely the modulations at the output of each filter/channel of the filter bank that constitutes the auditory periphery, these energy fluctuations can be characterized by the "modulation spectrum" (Greenberg 2005; Kanedera et al. 1999). For speech produced at a natural rate, the modulation spectrum across languages peaks between 4–6 Hz (e.g., Elliott and Theunissen 2009). Critically, these energy modulations correspond in time roughly to the syllabic structure (or syllabic "chunking") of speech. The syllabic structure, as reflected by the energy envelope over time, in turn, is perceptually critical because (a) it signals the speaking rate, (b) it carries stress and tonal contrasts, and (c) cross-linguistically the syllable can be viewed as the carrier of the linguistic (question, statement, etc.) or affective (happy, sad, etc.) prosody of an utterance. As a consequence, a special sensitivity to envelope structure and envelope dynamics is critical for successful auditory speech perception.

One hypothesis about a potential mechanism for chunking speech (and other sounds) is based on the existent neuronal infrastructure for dealing with temporal processing in general. In particular, *cortical oscillations* could be efficient instruments of auditory cortex output discretization/chunking/sampling. Neuronal oscillations reflect synchronous activity of neuronal assemblies (either intrinsically coupled or coupled by a common input). Importantly, cortical oscillations are argued to shape and modulate neuronal spiking by imposing phases of high and low neuronal excitability (e.g., Fries 2005; Schroeder et al. 2008). The assumption that oscillations cause spiking to be temporally clustered derives from the observation that spiking tends to occur in the troughs of oscillatory activity (Womelsdorf et al. 2007). It is also assumed that spiking and oscillations do not reflect the same aspect of information processing. While spiking reflects axonal activity, oscillations are said to reflect mostly dendritic postsynaptic activity (Wang et al. 2012).

Neuronal oscillations are ubiquitous in cerebral cortex and other brain regions (e.g., hippocampus), but they vary in strength and frequency depending on their location as well as the exact nature of their generators. In human auditory cortex, at rest (i.e., no input), ~40 Hz activity (low gamma band activity) can be detected (using concurrent EEG and fMRI) in the medial part of Heschl's gyrus, a region that is situated just next to core primary auditory cortex. In response to linguistic input, gamma oscillations spread to the whole auditory cortex as well as to classical language regions, where they cannot be detected at rest (Morillon et al. 2010).

If there exists a principled relation between the temporal properties of neuronal oscillations and the temporal properties of speech (i.e., delta band/intonation contour, theta band/syllabic rate, gamma band/segmental modulation), it stands to reason that these correspondences are not accidental. The speech processing system is exploiting the neuronal, biophysical infrastructure and yielding speech phenomena at timescales provided. In this context, it is worth remembering that the observed neuronal oscillations are not merely "driven in" to the system by external signal properties but are

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

rather endogenous aspects of brain activity. Indeed, experimental data from many animal studies as well as some recent human data show that neuronal oscillations in these ranges are endogenous and evident in auditory and motor areas (Giraud et al. 2007; Morillon et al. 2010).

Such data suggest an intriguing evolutionary scenario in which neuronal processing timescales follow from purely biophysical constraints (and therefore will also be visible in other primates) for the basis for timing phenomena in speech processing. The cognitive system is grafted on top of structures that provide hardware constraints, setting the stage for potential coevolutionary scenarios of brain and speech. (Fogassi, this volume, offers a complementary perspective on the evolution of speech; Arbib and Iriki, this volume, place more emphasis on the role of gesture in the evolution of the language-ready brain.)

Final Remarks

The data from neurobiology, cognitive neuroscience, psycholinguistics, and linguistics lead to a similar conclusion across domains: there is no single computational entity called "syntax" and no unstructured operation called "semantics," just as there is no single brain area for words or sounds. Because these are structured domains with considerable internal complexity, unification, or linking operations as outlined in the MUC perspective above, is necessary. Cognitive science research, in particular linguistic and psycholinguistic research, shows convincingly that these domains of processing are collections of computational subroutines. Therefore it is not surprising that the functional anatomy is not a one-to-one mapping from putative language operation to parts of brain. In short, there is no straightforward mapping from syntax to brain area X, semantics to brain area Y, phonology to brain areas Z, etc. Just as cognitive science research reveals complexity and structure, so the neurobiological research reveals fractionated, complex, and distributed anatomical organization. Moreover, this fractionation is not just in space (anatomy) but also in time: different computational subroutines act at different points in the time course of language processing. When processing a spoken sentence, multiple operations occur simultaneously at multiple timescales and, unsurprisingly, many brain areas are implicated in supporting these concurrent operations. The brain mechanisms that form the basis for the representation and processing of language are fractionated both in space and in time, necessitating theories of unification that underpin how we use language to arrive at putatively unified interpretations.

Music is in many ways like language. Although it is not very helpful to try to make direct comparisons between building blocks of music and language (e.g., to claim that words correspond to notes), music is almost certainly another complex faculty that has to be decomposed in multiple subroutines, each recruiting different nodes in a complex neuronal network. It is likely

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.

that some of the nodes in the neuronal networks that support the perception and production of music are shared with language. In both cases, meticulous analyses is required to determine what the primitives are (for a discussion of this approach and an attempt to make explicit what is shared and what is different, see Fritz et al., this volume); that is, what the "parts list" is (e.g., features, segments, phonemes, syllables, notes, motifs, intervals). This will enable us to meet the challenge of mapping the list of primitives for language and music to the computations executed in the appropriate brain areas.

From "Language, Music, and the Brain," edited by Michael A. Arbib.
2013. Strüngmann Forum Reports, vol. 10, J. Lupp, series ed. Cambridge, MA: MIT Press. 978-0-262-01810-4.