7 Key Issues and Future Directions: Toward a Comprehensive Cognitive Architecture for Language Use

JAMES M. MCQUEEN AND ANTJE S. MEYER

Language can be used in many different ways. We can, for instance, have a chat with a neighbor, listen to a lecture, write a novel or a manual, cite a poem, or read a scientific text or newspaper. All of these activities involve components of the cognitive architecture that supports linguistic communication. The preceding five chapters in this section discussed important issues that must be addressed in theories of this architecture: the nature of linguistic representations (Jackendoff, chapter 2), the processes involved in understanding spoken language (Dahan & Ferreira, chapter 3) and written language (Andrews & Reichle, chapter 5), those involved in speaking (Roelofs & Ferreira, chapter 4), and the cognitive implications of the multimodal nature of language (Özyürek & Woll, chapter 6). In this chapter, we adopt a broader perspective. We propose a research program to work toward integrative models that account for all of the ways that language is used. Such models should explain speaking, listening, reading, and signing in a coherent and comprehensive way. We discuss which design choices have to be made in generating such models, and how assumptions about the broad architecture of a model and about its components constrain each other.

To start, we categorize language tasks—all the things that people can do with language—by considering the language user's current goal, namely to understand or to produce language, and the modality, namely spoken, written, or signed language. This broad categorization of language tasks and the ways different tasks are related to each other are further discussed in section 1.

We assume that all language tasks involve two types of cognitive components: (i) linguistic and nonlinguistic representations, and (ii) cognitive processes that retrieve or operate on these representations in the order appropriate for the task. Following the principle of parsimony, research should work toward finding the simplest possible characterization of these components. The goal of our proposed research program is thus to describe a single set of representations and processes required for all linguistic tasks and to specify which combination is used and in what way in each task. The set should be as small as possible but sufficient to accomplish all tasks in all languages.

Our working hypothesis implies that there is one language system, consisting of multipurpose linguistic representations (see section 2) and multipurpose procedures to access linguistic representations and perform operations on them (see section 3). Likewise, we propose one domain-general cognitive system (see section 5) that is implicated in linguistic and nonlinguistic tasks. These assumptions do not imply that exactly the same representations or processes are involved in exactly the same way in all tasks; for instance, access to orthographic representations undoubtedly plays a central role in reading but is less important in speaking. Our proposal simply holds that there is one knowledge base and one set of processing mechanisms that are recruited in different ways depending on the specific language task being performed. Some representations or processes may well be unique to a given task.

This characterization of the cognitive architecture for language may suggest that it is a facility that is used over and over again in the same fashion, but in fact a person's language architecture changes with experience across the life span. As discussed in section 4, this plasticity needs to be captured in any theory, as learning and adaptation play crucial roles in many linguistic tasks and establish links between them, with, for instance, learning to comprehend a word being a precursor for someone being able to produce that word.

Our proposal raises three closely related challenges for future research. The first is to specify the minimum representational and processing requirements, that is, to work out which representations and processes are required to perform all language tasks. The second challenge is to delineate which representations and processes are used in each individual task and hence to establish the simplest possible comprehensive model that is consistent with the data pertaining to the task. Third, it needs to be specified how specific processes and representations are combined and ordered in each task.

1. Relationships among Language Tasks

1.1. Producing and Comprehending Language

Our proposed research program starts from the assumption that language production and comprehension are different sets of tasks rather than distinct processing systems. Clarifying the relationship between these tasks in terms of shared or unique cognitive processes will be a major step toward developing a comprehensive model of language use.

The tasks of speaking and listening, or producing and understanding sign language, are often said to differ in the direction of the information flow, from concepts to articulation versus from phonetic input to concepts. But comprehending is not speaking (or signing) in reverse. The cognitive challenges arising for speakers and listeners are fundamentally different: Speakers start from a conceptual structure, select and order units and eventually produce overt behavior, which they monitor for appropriateness and correctness (e.g., Levelt, 1989). Listeners do not have to react overtly (and hence probably do not monitor their behavior to the same extent), and they do not have to order units, as the order is provided by the speaker. Their task is to segment the continuous speech stream into words, find the syntactic structure and grasp the speaker's meaning (e.g., Cutler & Clifton, 1999). A main challenge for listeners is to deal with ambiguities arising at all levels of processing (cf. the noisy channel model discussed by Dahan & Ferreira, chapter 3) and to do so at the speed set by the speaker.

Another frequent characterization of the relationship between the tasks of comprehension and production is that they both involve two systems (a production system and a comprehension system). It has long been acknowledged that speakers use a speech production system but also listen to their own overt and inner speech using a speech comprehension system and may use this system to monitor and correct their utterances (e.g. Levelt, 1989; see also Roelofs & Ferreira, chapter 4). In the speech perception literature, it has often been argued that speech recognition depends on the speech production system (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). In addition, many authors have recently argued for the involvement of the production system in higher-level speech comprehension. Specifically, there is strong evidence for the importance of prediction of upcoming content during reading and listening, and making predictions is often assumed to involve processes that play a central role in production (Pickering & Garrod, 2013). These proposals, that speakers can listen to themselves and that they use their language production system to predict what others might say, fit our experience as speakers and listeners. However, for generating a comprehensive model of language use, they are not particularly helpful since they start from the assumption that there are distinct comprehension and production systems, which nevertheless are both involved in comprehension and production tasks. The proposal that we advocate here does not postulate distinct systems.

We propose that research toward comprehensive models should instead presuppose a single language system. This research then needs to specify which knowledge structures and which processes are implicated in specific production and comprehension tasks and, following on from this, which components of the language system are widely shared between tasks and which are unique to some tasks. This requires conducting research projects that simultaneously consider specific production and comprehension tasks, for instance, word production and word comprehension. In such projects, one might find that some production and comprehension tasks are more closely related to each other than some pairs of tasks within the production family or within the comprehension family.

One important theoretical and empirical challenge in such a research program is to be clear about what it means for a component to be shared across tasks. In a general sense, all language tasks draw on shared conceptual and linguistic knowledge. Though this is rarely discussed, the consensus in the field appears to be that the same conceptual stratum is accessed in production and comprehension. Linguistic knowledge must also be shared since we can only learn to speak a language by hearing it, or to sign a language by seeing it. However, a much debated issue is whether comprehension and production draw on the same representations or whether there are dedicated, but tightly linked representations for the two types of tasks (cf. Meyer & Huettig, 2016). The same-representations view is parsimonious, but it leads to challenging questions about the processes involved in production and comprehension. Most obviously, how do the production deficits arise that are so prominent in children and second-language speakers of a language? And why are estimates of receptive vocabulary usually larger than those of productive vocabulary? The separate-representations view faces a different challenge, namely to explain how task-specific representations are linked such that speakers learn to produce utterances from hearing them.

Studies comparing production and comprehension must zoom in on particular linguistic levels or processing components, such as grammatical, morphological, or phonological encoding of sentences. For each level, we would expect to find the engagement of both shared and unique components. For instance, grammatical encoding processes must occur in production and in comprehension. Perhaps the underlying processes are the same, maybe consisting of combining treelets as proposed by Jackendoff in chapter 2. However, any shared processes must co-occur with unique processes, since grammatical encoding during speaking is driven by the speaker's conceptual representations, which are not usually ambiguous, whereas grammatical encoding during comprehension is driven by the speech input, which is riddled with (at least temporary) ambiguity. Similarly, phonetic and phonological encoding and decoding of utterances may involve shared word-form representations, but phonological encoding during speaking involves serial ordering of segments and preparation of articulatory commands, whereas phonological encoding during listening involves the parsing of a speech signal into words. Ambiguity and ordering requirements may be the main causes of differences in the processes involved in different tasks. Thus, a research program that seeks to establish which representations and processes are used in which tasks may uncover general principles that determine similarities and differences in the processes subserving different tasks.

This discussion has concerned speaking and listening, but similar considerations apply to producing and understanding signed languages and to reading and writing. In all cases, the tasks are different, but are likely to draw on shared as well as unique representations and processes. Functional models of the language system must specify which representations are shared, what it means to share representations, and which processes occur in several tasks or are specific to some of them.

1.2. Using Language in Different Modalities

The modality—spoken, written, or signed language—obviously affects how the linguistic tasks of producing and comprehending language are carried out. Readers only have visual information to rely on; listeners sometimes, for instance, in telephone conversations, have only auditory information to rely on, whereas interlocutors in natural conversations, as Özyürek & Woll (chapter 6) highlight, typically process complex multimodal information streams. Similarly, writers typically use hand movements to produce written output, whereas speakers use their speech motor system and often their hands and body to gesture. The involvement of different sensory and motor systems in different tasks suggests that many of the cognitive components involved in these tasks must be different as well.

In addition to processing differences directly related to the use of the auditory or visual modality, there are processing differences that stem from specific properties of spoken or written language. Many of them are discussed by Dahan and Ferreira (chapter 3) and Andrews and Reichle (chapter 5). For instance, listeners have to segment the speech stream into words and they have to cope with speaker variability as well as disfluencies and errors. The reader's task seems easier, as printed words in many writing systems are separated by blanks, there is limited variability across type fonts (though not in handwriting), and most texts are edited and probably contain fewer errors than spontaneous speech. In contrast, compared to the complex speech input available in everyday conversation, written language is impoverished; there is, for instance, no prosodic information, and there are no gestures and facial expressions clarifying speaker meaning.

In spite of these differences, the knowledge structures and processes involved in using written and spoken language are tightly related. As Andrews and Reichle discuss, children learn to read by linking written words to spoken ones. Moreover, their model of adult reading includes the claim that readers map written forms onto spoken word representations and that higher-level text comprehension processes use largely the same mechanisms as are used in spoken language comprehension. Similarly, the mechanisms involved in conceptual and grammatical encoding are probably shared between speaking and writing.

In short, parallel language tasks carried out in different modalities (e.g., recognizing written and spoken words) undoubtedly involve some shared and some unique cognitive representations and processes. An important theoretical issue is then what it means for representations or processes to be shared. For instance, do speakers retrieve the same word-form representations for speaking and writing and then activate separate output representations, or do they retrieve dedicated modality-specific representations? It needs to be established how tightly linguistic representations are linked to specific input or output modalities (see also section 2.3).

Another important issue is to what extent and how language input or output in different modalities requires the use of distinct domain-general processes. For instance, both reading and listening require rapid incremental processing with an occasional need to revise structures or interpretations generated earlier.

As Andrews and Reichle point out, the mechanisms involved in combining word meanings and generating syntactic structures are likely to be shared across modalities. But the demands on domain-general processes are different. Listeners need to process sentences at the pace set by the speaker and have to base any revisions on working memory representations of earlier input. By contrast, readers can set their own pace for reading, and when reanalysis is necessary they can redirect their gaze to earlier sections of the text. Thus, the attentional demands of reanalyzing spoken and written sentences are different, involving memory retrieval of earlier information or redirecting visual attention to earlier text, respectively. How these differences affect the generation of the grammatical and semantic structure of sentences is largely unknown.

Sign language is a third modality, in addition to spoken and written language. However, the relationship between signing and the other two modalities is different from the relationship between the spoken and written versions of a single oral language (e.g., English) because readers and writers of an oral language can draw on representations that are also implicated in their spoken language, whereas for signers the signed and written languages are different. For instance, users of Dutch Sign Language may be readers of Dutch and English. Thus, the question of whether or not shared representations of specific words or grammatical rules are activated in a person's mind in signing and listening, or signing and reading, does not necessarily arise (mouthing, e.g., concurrent articulation of English words during signing in British Sign Language is an interesting exception). As Özyürek and Woll (chapter 6) demonstrate, however, one can still ask which characteristics of linguistic representations underlying spoken and signed languages are functionally equivalent, which processes occur in both modalities, and which specific constraints arise in each modality.

2. How Is Knowledge about Language Stored in the Mind?

If, as we propose, language production and comprehension are seen as different language tasks (rather than distinct language systems), an important goal for comprehensive models of language is to specify how the tasks are related in terms of the domain-general and linguistic representations and processes they involve. In the simplest model, all language tasks are supported by shared memory representations of language. The single-store assumption is more parsimonious than assuming separate representations for separate tasks. Moreover, the single-store assumption is plausible since, at least for monolingual speakers, there is only one language to represent and there is only one brain and mind to represent it. Whether the assumption is correct is an empirical issue. Future research will also have to adjudicate on several design choices about the representation of linguistic knowledge.

2.1. The Contribution of Linguistic Theory Evidently any psychological theory of language knowledge must be informed by linguistic theory; even the most common descriptive terms, such as words and phrases, are linguistic terms. It is less clear, however, how much linguistic detail needs to be included in cognitive models of the language system. Jackendoff (chapter 2) makes a strong plea for linguistically sophisticated representations: interdisciplinary language science, he argues, would make a fundamental mistake if it ignored the complexity of linguistic structure. In contrast, the other authors in this section present models with relatively simple representations, where broad concepts such as *letters*, *words*, and *sentences* do much of the work.

Psycholinguistic models often fail to consider many aspects of language that quite likely affect speakers' and listeners' behavior. For instance, while there is a substantial body of psycholinguistic work on the representation of the morphological structure of words (for a recent review, see Taft, 2015), there is, as Dahan and Ferreira (chapter 3) point out, still far too little work on prosodic representations. Moreover, psycholinguistic research has largely concentrated on a number of closely related languages. An ambitious goal for a comprehensive model of language use is to account for the ways language tasks are accomplished in all (or at least a variety of structurally different) languages. Sophisticated linguistic descriptions of different languages would be extremely valuable (e.g., by indicating which languages need to be tested) in achieving this goal.

Even though models of language use must include knowledge of linguistic structure, there need not be a one-to-one correspondence between the structures used in accounts of language itself and those used in accounts of language-based communication. This is because the explananda of the two disciplines (language structure and language cognition) are different. Even though much linguistic theorizing is closely tied to claims about cognition and/or is often based on performance data, linguistic theory does not equate with cognitive theory. There can therefore be misalignments between linguistic descriptions and language users' behavior. For instance, while linguistic theory might consider a particular word to be morphologically complex, users may consider it to be morphologically simple (Marslen-Wilson, Tyler, Waksler, & Older, 1994).

In our view, linguistic theory is indispensable for providing hypotheses about the way language might be represented in the brain and mind. These hypotheses need to be empirically evaluated. A parsimonious cognitive model should include those, and only those, aspects of linguistic representation that demonstrably affect how people solve linguistic tasks.

2.2. Lexicon and Grammar An important design question for models of language use is whether and how to represent the classic distinction between the lexicon and the grammar. Knowledge about language is likely to be stored in long-term memory in part in a declarative way (e.g., through representations that capture knowledge of linguistic structures) and in part in a procedural way (e.g., through processes that use or act on those representations). Declarative knowledge about the meaning of words, for example, may interface with procedural knowledge about word-order rules. A straightforward way of realizing the distinction between grammar and lexicon in psychological models of language has been to equate the lexicon with declarative knowledge, and the grammar (including syntactic, semantic, and phonological rules) with procedural knowledge.

Unfortunately, there is no simple theoretical divide between declarative lexicon and procedural grammar. A standard position is that all rule-based knowledge is grammatical, while exceptions are stored in the lexicon. As Jackendoff (chapter 2) discusses, however, knowledge about regular syntactic constructions can be stored in the lexicon. Grammatical knowledge then becomes declarative rather than procedural. Similarly, empirical evidence is not easy to categorize: In many cases, a given finding could arise either because the language user has stored some knowledge in a representation or because that knowledge is instantiated in a given procedure. Evidence of sensitivity to the phonotactics of a language, for example, could reflect lexical storage (legal sequences of segments appear in words, illegal sequences do not), or it could reflect the operation of phonological encoding or decoding mechanisms sensitive to the sequential probabilities of segments. Cognitive theories therefore disagree about where to draw the line between declarative and procedural knowledge (see, e.g., Christiansen & Chater, 2016, and accompanying commentaries).

Cognitive theories disagree not only about storage versus computation but also about whether there is a distinction between the representations that are used for online processing and those that are stored in long-term memory. For instance, the word recognition model Shortlist (Norris & McQueen, 2008) distinguishes between the representations used in processing and those held in the lexicon. There is no such distinction in TRACE (McClelland & Elman, 1986): word nodes in long-term memory vary in activation level to represent lexical hypotheses about the current speech input. As theorists attempt to build more comprehensive models, they will therefore need to specify not only what the balance is between computation and storage, but also, at the interface between the two, whether there is a distinction between the temporary structures representing the here-and-now of current processing and the representations in the long-term store.

2.3. Abstract versus Embodied Representations of Word Meaning A third important design question concerns the representation of word meaning. Specifically, should word meanings be considered as abstract symbolic representations, which are linked to nonlinguistic sensory and motor representations of the same concepts, or is language fundamentally embodied so that a categorical distinction between abstract linguistic and nonlinguistic representations is incorrect? The authors of the current section adopt different views on this highly contentious issue: Roelofs and Ferreira (chapter 4) assume abstract symbolic linguistic representations that are linked to supramodal conceptual representations, which in turn are linked to modality-specific conceptual features. In contrast, Özyürek & Woll (chapter 6) argue that language use is typically multimodal (involving, for instance, speech and gesture), and that language representations likely consist not only of discrete and arbitrary abstract symbols but also of analog and nonarbitrary components (e.g., iconic components in co-speech gestures).

More work is required to determine the best way of representing word meanings. The available evidence indicates that words can rapidly evoke related sensory information and motor programs (e.g., Barsalou, 2008; Pulvermüller, 2005). It is not yet clear, however, whether sensory and/or motor activation are mandatory components of word processing, that is, whether they always arise or only in specific tasks and contexts. Evidence for mandatory sensory or motor activation would strengthen the case for integrated multimodal representations of word meanings, whereas evidence for context- or task-dependent activation would argue for abstract representations with links to sensory and motor components. Note that here, and in many other research contexts, considering several language tasks (for instance word production and reading) simultaneously may contribute to solving the question at hand.

Important theoretical issues concern the best ways of formalizing representations of word meanings.

Regardless of whether word meanings are seen as intrinsically multimodal or as abstract representations with links to motor and sensory representations, ways need to be found to represent these components of knowledge in compatible ways such that mutual influences between them can be captured. It also remains to be seen which representational format, for instance, symbolic or subsymbolic (see, respectively, Page, 2000, and Seidenberg & McClelland, 1989, for examples), is most suitable for linguistic knowledge.

2.4. Abstract versus Episodic Representations of Word Form A related key dichotomy is whether knowledge about the forms of words is abstract or episodic in nature. That is, does the representation of the form of a word include episodic details of, for example, the way a particular speaker said that word on a particular occasion? As Dahan and Ferreira (chapter 3) discuss, talker-specific detail appears to help listeners tune in to and hence better understand their interlocutors. But these talker-specific details appear to modulate a process of phonological abstraction (McQueen, Cutler, & Norris, 2006) and thus must also be stored. There is a growing consensus that a hybrid model for form representation is required, where episodic storage (e.g., about talker idiosyncrasies) is combined with cognitive abstraction.

The hybrid model needs to be specified in more detail. How are episodic and abstractionist components stored and how are they combined in online processing? Adaptive ideal observer models (Kleinschmidt & Jaeger, 2015) have recently offered an interesting approach to this question for speech perception: Listeners are considered to make inferences about uncertain input based on distributional knowledge and to adapt when faced with novel input (e.g., from a new talker). Answers, however, need to be provided for other levels of processing and across modalities and tasks. For example, what is the balance between abstract and episodic memory in speech production? (See Pierrehumbert, 2002, for one suggestion.) From the perspective of our proposed research program, the question that therefore needs to be addressed is whether the hybrid use of abstract and episodic representations of word form that appears necessary for speech comprehension is also required in other language tasks.

3. What Are the Component Processes in the Cognitive Architecture?

According to our proposal, different combinations of processing mechanisms are called on in different language tasks. In designing comprehensive models of the language system, design choices need to be made about the broad partitioning of the system into processing stages and the time course and direction of the information flow between them. In addition, the processing mechanisms themselves need to be specified.

3.1. How Many Distinct Processing Stages? The first design choice about the general architecture of the language system is whether or not there are any distinct processing stages. That is, are there processors that handle different types of information and which operate separately, in terms of either time (i.e., one processor begins before another) or information exchange (i.e., one processor does not influence the operation of another)? Compare, for example, the model of speaking presented by Roelofs and Ferreira in chapter 4 with that for listening by Dahan and Ferreira in chapter 3. There is much greater compartmentalization in the production model. Roelofs and Ferreira distinguish between different components of speaking (conceptualization, lexical access, and articulatory encoding), while Dahan and Ferreira argue for a model in which knowledge sources at multiple levels of representation (e.g., phonetic, lexical, semantic) jointly determine comprehension. The comprehension model does not deny that different levels of representation are involved; the claim is that there are no clearly separable stages of processing. Which view on the compartmentalization of processing is correct, or is it the case that speaking and listening are fundamentally different?

This debate about processing stages contrasts modularity (e.g., Fodor, 1983) with interactivity (e.g., McClelland & Rumelhart, 1981). It concerns different degrees of granularity of the stages. That is, it includes issues about the extent to which the language system as a whole is separable from other domains of cognition (see section 5), whether the processes and representations subserving different language tasks are separable (see section 1), and whether different components responsible for a given task or subtask are separable. For a complete model of language use to emerge, researchers need to collect data that specify which processes interact and which do not. This is not trivial because data showing interactive effects of two variables (for instance a conceptual and a syntactic variable) on a speaker's or listener's behavior are not sufficient to show that underlying processes interact. For instance, semantic and syntactic processes can be independent but the output from those two processes can still jointly determine a decision. What is required to demonstrate an interaction of two processes are data showing that one process influences the operation of the other process (see Norris, McQueen, & Cutler, 2016, for further discussion).

Many psycholinguistic models take some parts of the system for granted. For example, older models of parsing such as the sausage machine model (Frazier & Fodor, 1978) took strings of words that had already been recognized as input for grammatical processing. Such models thus ignored the possibility that delays in the recognition of the phonological form of words (e.g., due to temporary perceptual ambiguities) could interfere with parsing and meaning construction. In contrast, the constraint-based account proposed by Dahan and Ferreira (chapter 3), in which form, syntax, and meaning jointly constrain comprehension, allows for the possibility of interactions between phonological and grammatical processing. In the future, experimental and computational work aiming at a comprehensive model will need to focus in particular on the interfaces between the main types of processes (such as the interface between form and syntax) to establish whether different stages of processing are indeed separable.

In addition to the question about how many distinct stages of processing may be involved in a given language task, there are two interrelated questions about the way the different stages, if they exist, talk to each other. First, is processing serial or cascaded? Second, is processing unidirectional or bidirectional? At one extreme, if processing goes through fully discrete and serial stages, there is no cascade (i.e., no continuous feedforward flow of information) and no bidirectionality (i.e., no feedback either). At the other extreme, there is cascaded flow of information in both directions. But intermediate accounts, that is, models with cascade but without feedback from one stage of processing to the next, are also possible.

3.2. Cascaded Processing There is consensus that, for both comprehension (Dahan & Ferreira, chapter 3) and production (Roelofs & Ferreira, chapter 4) of spoken language, processing is cascaded. A logically later stage of processing does not need to wait until an earlier stage has been completed; instead, information is passed continuously forward. Cascade of information makes it possible for processing to be, in temporal terms, incremental. For example, in comprehension, information concerning different levels of description (e.g., semantic and phonetic information) appears to be used by the listener as soon as it comes available (see Dahan & Ferreira). Thus, any comprehensive model of the language system is not likely one that assumes that there are only strictly serial stages of processing. However, as Andrews and Reichle (chapter 5) argue with respect to reading, there may, nonetheless, be serial steps in some aspects of language processing.

The consensus on cascaded processing within the speech production and perception literature is a good example of progress; 20 years ago this issue was still open (compare, for instance, cascaded models of speech production such as that of Dell, 1986, with serial models such as that of Levelt, Roelofs, & Meyer, 1999). We suggest that this issue was resolved for several reasons: because the issue was clearly formulated, because the question was simple, and, ultimately, because careful experimentation yielded converging evidence from multiple methods (see Roelofs & Ferreira, chapter 4, for discussion of this evidence for speech production).

3.3. Feedback? In contrast, the question about directionality of information flow has not yet been resolved. Most researchers would agree that extreme modularity (in which flow is strictly unidirectional in all parts of the language system) is untenable. For example, there is evidence of facilitatory effects of lexical neighborhood density in speech production (Vitevitch, 2002) that is difficult to reconcile with purely feedforward processing; in speech recognition, there is evidence that lexical knowledge can be used to retune prelexical processing of speech segments (Norris, McQueen, & Cutler, 2003; Dahan & Ferreira, chapter 3). Nevertheless, there are still many open issues about the limits and the nature of bidirectional processing. Consider, for example, studies on lexical involvement in phonetic categorization (Ganong, 1980). Here, profeedback conclusions (Elman & McClelland, 1988; Gow, Segawa, Ahlfors, & Lin, 2008) contrast starkly with antifeedback conclusions (Kingston, Levy, Rysling, & Staub, 2016; McQueen, Jesse, & Norris, 2009).

Why has this question proved so much harder to resolve than that on cascade? The issue seems clear, and a great deal of research has been done, yet no consensus has been reached. We suggest that the reason why this issue—in the speech perception domain and others (including reading; see Andrews & Reichle, chapter 5)—is still open is that it is not as simple as it appears to be. It is not sufficient to ask only whether or not processing is bidirectional. Instead, researchers need to define the nature of the processing under investigation. If there is feedback from one stage to a preceding stage, what function does it serve? Is it about current linguistic content, for learning or for attentional control, or does it act to bind together representations at different linguistic levels? Apparent contradictions in the literature might be resolved if it can be established that different types of feedback are involved in different experimental tasks and settings.

3.4. Shared Processing Mechanisms across Tasks In addition to describing the broad partitioning of the language system into processing stages and the information flow between them, models of the language system must specify the mechanisms that are involved in each task.

According to our proposal, the same mechanisms are available for all tasks. These would therefore need to be computational primitives that can apply to different types of representations. Various mechanisms that can apply throughout the language system have been proposed in the literature. For instance, all models we know of assume that in comprehension and production, across modalities, lexical items are accessed and selected. These operations are often defined in terms of activation of units and selection among them. Selection may or may not be a competitive process, and it may not involve inhibition. Models concerning the processing of larger units often assume operations that map lexical units onto slots in grammatical frames (Roelofs & Ferreira, chapter 4) or that merge or "clip together" (Jackendoff, chapter 2) successive utterance fragments. To give a final example, many contemporary models stress the importance of predictive processes for all linguistic tasks (Dahan & Ferreira, chapter 3).

An important theoretical and empirical challenge is to determine at which levels of granularity, or in which "vocabulary," processing mechanisms in integrated models of the language system can best be described. Should they be at the level of computational primitives (simple and broadly applicable) or combinations of those primitives (more complex and less broadly applicable)? Another empirical challenge is to demonstrate that the postulated processing mechanisms are indeed sufficient to characterize how people carry out the whole range of linguistic tasks. The history of the field suggests that everyday terms such as *activation*, *selection*, and *inhibition*, which can be transparently implemented in computational models, are the most useful. Parsimony dictates that the processing mechanisms should be the simplest and the most broadly applicable.

Processing mechanisms are likely to differ in their applicability across tasks. For instance, there may be mechanisms that are exclusively involved in segmenting the speech stream and do not play a role in other tasks. Likewise, there may be mechanisms that are of particular importance in processing some languages, but only play a minor role in others (see, e.g., El Aissati, McQueen, & Cutler, 2012, who summarize research on speech segmentation across 10 typologically diverse languages). A key issue to be addressed is which features of the cognitive architecture of the language system are shared across languages and which depend on characteristics of the particular language. In other words, the issue is which, if any, processing universals (rather than linguistic universals, cf. Evans & Levinson, 2009) can be identified.

4. Language Use and Language Learning

In this chapter, we (and our coauthors in this part of the book) focus on language use and do not consider extensively the process of first- or second-language acquisition. However, language use and learning are inextricably bound up with each other. Infants and children learn spoken language in order to be able to use it. Reading is a skill taught and practiced in primary-school classrooms. There is a growing body of evidence that the processing components of the major language tasks are not static but change with language use. Readers, speakers, and listeners are flexible (Dahan & Ferreira, chapter 3). This flexibility also transfers across tasks, such that, for example, learning in speaking can be achieved through listening (Kittredge & Dell, 2016).

While models of reading often include an acquisition component (see Andrews & Reichle, chapter 5), models of speech comprehension and speech production for many years tended to be static. They offered accounts of particular aspects of speech processing but did not capture learning or adaptation in any way. Things are changing, however (see, for instance, Christiansen & Chater, 2016; Dell & Chang, 2014; Kleinschmidt & Jaeger, 2015). A challenge for the development of comprehensive models of the language system will be to further specify the balance that must exist between stability and flexibility. Clearly the processing system adapts in the light of new experience, but it must do so in a way that maintains stable comprehension and production abilities. Another challenge will be to clarify the relative contributions of experience (e.g., hearing a talker speaking a different dialect; cf. Dahan & Ferreira, chapter 3) and maturational change (how the brain and mind change through development across the life span) to these dynamic accounts of language processing. Since learning often involves several language tasks (for instance listening and speaking or listening and reading), work on language learning will contribute to a better understanding of the relationships between these tasks.

As this work advances, it is likely that the traditional boundary between language acquisition and language processing will be further eroded. Ultimately, the goal of language learning is not to acquire linguistic knowledge (e.g., vocabulary, grammatical rules, pronunciation), but to be able to speak and listen and read and write and sign. Second-language learners may sometimes achieve this goal with inadequate or incomplete knowledge but even for first-language learners, the ultimate goal is communication, not the acquisition of knowledge alone. In our proposed approach to the cognitive architecture, learning and adaptation are key processes. From this perspective, however, learning should be seen as supporting language use rather than as passive knowledge acquisition.

5. The Relationship of the Language System to Other Components of the Cognitive System

As we have discussed, language tasks draw on motor and sensory components that are also involved in nonlinguistic tasks. Comprehensive functional models of the language system must spell out how the language system is related to these other components of the cognitive system. Thus, models must specify how linguistic representations are related to sensory and motor representations (see section 2.3) and how the processing requirements of specific tasks (e.g., the need to process auditory or visual information) impinge on linguistic processes.

In addition, language tasks draw on domain-general attention and memory processes. The importance of these systems for language processing has long been recognized. The chapters in the present section discuss the involvement of the domain-general cognitive system in language processing in several ways. For instance, the role of visual attention for language processing is discussed by Roelofs and Ferreira (chapter 4) in their section 4.3 and by Andrews and Reichle (chapter 5) in their section 3.1. Andrews and Reichle highlight the importance of visual attention in linearizing texts, thereby turning the simultaneously present visual information into a temporally ordered sequence of chunks that can be further processed in a working memory buffer and by higher-order processes, in ways very similar to those involved in processing spoken information. The importance of working memory for language processing is also stressed by Roelofs and Ferreira (chapter 4), by Andrews and Reichle (chapter 5), and by Dahan and Ferreira (chapter 3).

A comprehensive model of the language system must indicate how and where specific linguistic processes and domain-general central cognitive processes interact. Currently, the field is fractionated, as researchers have related specific components of the language system (e.g., auditory word processing or comprehension of written sentences) to specific components of attention or memory systems (e.g., inhibitory control or working memory). The components of attention and memory systems are defined in different ways in different studies, and the relationships between these components are far from clear. This state of affairs makes it difficult to determine exactly how attention and memory are involved in different linguistic tasks.

Thus, in addition to design decisions about the language system, decisions are required about the properties of all of the other cognitive systems involved in language tasks. Making such decisions is daunting but not impossible. For instance, Roelofs and collaborators (see references cited in Roelofs & Ferreira, chapter 4, section 4.3) adopted a specific theory of the attentional system (following Posner, 2012, and Miyake et al., 2000) and investigated systematically how different attentional components were involved in different steps in the process of word production. This led to the development of a clearly articulated theory of the involvement of attention in word production. It should be possible to extend this approach to other linguistic tasks. In a comprehensive model of the language system, the involvement of domain-general processes must be clearly specified. In such a model, the smallest possible set of domain-general processes should be implicated, in different combinations in all language tasks.

6. Conclusion

The major challenge for current psycholinguistics is to build a comprehensive theory of all aspects of cognition that supports linguistic behavior. We have taken the stance that there is a single system supporting all language use, rather than separate systems for different language tasks (writing, speaking, signing, and comprehending text, speech, and sign). The theory of this system contains the smallest possible set of cognitive representations and processes that the language user needs to carry out all linguistic tasks and specifies how these representations and processes are combined in each task.

In our proposal, the knowledge base is the same across tasks. It includes a lexicon, with word-form and word-meaning representations, stores containing the sensory and motor components of language, and a grammar (capturing not only syntactic rules but also, e.g., phonological rules). It is unlikely that all components of the knowledge base are used in all tasks. For instance, knowledge about the acoustic properties of a vowel is likely to have little role to play in sign production. Nevertheless, our proposed research strategy is to specify, first, what the minimal set of representations and processes are required for all tasks, and second, which of them are used in which task. As discussed in section 2, it will be necessary to make many other design choices about the knowledge base: to specify how complex linguistic representations are, to delineate which aspects of knowledge are declarative and which are procedural, to clarify the extent to which representations of the meanings of words are or are not distinct from sensory and motor representations of the same concept, and to find the balance in the representation of word form between phonological abstraction and memory for episodic detail.

According to our proposal, the same processes are used across all language tasks and are built from a limited set of basic operations. The set of computational primitives may include retrieval processes (e.g., activation), selection processes (e.g., competition), slot-insertion processes, learning and adaptation processes, and predictive processes. Our proposed research strategy is, once again, to specify the minimal set of processes required for all language tasks and to determine which of them are used in which tasks. It will also be necessary to specify for each how the processes are combined. Are there distinct stages of processing and, if so, how distinct are they, both with respect to the timing of the operation and flow of information?

We have provided an overview of the issues that would need to be addressed in developing a comprehensive model of language use. Many of these issues are quite old. We think that it would be wrong to see this as an indication that understanding of the cognitive architecture of language has failed to advance significantly. Substantial advances have been made and the reason why many of the old questions still need to be answered simply reflects the fact that the questions are hard. Psycholinguists need to keep on chipping away at these hard questions so that our understanding of the cognitive architecture can cumulate.

We have chosen to center our discussion on the issues that will have to be addressed in order to generate a comprehensive model of the language system rather than on the methods that will be used to address them. There are, however, a few methodological points we would like to make. First, developing such a model requires the use of a wide range of research tools including linguistic analysis of language structures, behavioral and neurobiological experiments, and computational modeling. Such work is best done in interdisciplinary teams of scientists. Second, we encourage researchers to use replication and meta-analysis more extensively. This is for a very obvious reason: it needs to be clear what the reproducible core findings are; only then can we build a comprehensive theory of the language system.

Third, though the research goal is to develop the simplest possible model, much can be gained by considering variability between languages, between individuals, and between communicative situations. Considering variation between languages is crucial for discovering processing principles that hold across a wide range of languages and principles that are specific for certain types of languages. Exploiting variability between individuals is essential for correlational studies that relate, for instance, attentional skills to performance in linguistic tasks. Considering variability across communicative situations is crucial for understanding how people adapt to and learn from each other. In order to use and understand variability, new research tools may need to be developed, in particular for studying language use in everyday environments. Ultimately, psycholinguistics should explain not only how people use language in the solitary setting of a traditional laboratory task such as visual lexical decision but also, actually primarily, how they use it in natural settings. While we believe that experimental control is essential, and that research using laboratory tasks has revealed a great deal about normal language processing, techniques that allow for the investigation of language processing in more naturalistic situations such as two-person conversations can offer valuable complementary evidence.

Fourth, we expect that methods will change in the coming decades due to technological advances. Cognitive neuroscience techniques in particular have developed enormously in the last 25 years and will continue to do so. We similarly expect further advances in data collection and analysis that will facilitate much larger-scale investigations than were feasible even a few years ago. The researcher in the 1980s, for example, could barely imagine data collection from thousands of individuals using internet and mobile-device technology. We cannot imagine what will be possible in 2050. What we can say, however, is that whatever the technology used to do so, key questions about the cognitive architecture of language still need to be answered.

We recommend that researchers take a Marrian perspective to answer these questions. The analysis that Marr (1982) offered for vision, in which it could be considered at three levels, the computational, the algorithmic, and the implementational, applies equally well to language processing. The cognitive architecture of language is concerned more with the first two levels (the computational problems that the language user has to solve and the algorithms they use to do so) than with the third level (the neurobiological implementation). To take the Marrian perspective, therefore, is to start at the highest level: what are the computational problems that the listener or reader or signer or speaker has to solve? This question leads naturally to the algorithmic level: what representations and processes have to be in the language system to allow the language user to solve those problems? This perspective is very different from one that might be taken by linguists, in which the nature and structure of language is taken as the starting point rather than the cognition of the language user.

A Marrian perspective is also valuable when considering how the cognitive psychology of language relates to the neurobiology of language. One problem becomes immediately apparent. It is that there are multiple possible implementations for a given algorithm. This can make it very hard to draw strong inferences across levels of analysis, because a given algorithm does not demand a particular implementation, nor the reverse. Nevertheless, there are constraints, and these can work in both directions. That is, cognitive research can lead to proposals for possible implementations, and neurobiological research can indicate which kinds of cognitive algorithms are most plausible.

The Marrian perspective encourages a functional approach to neurobiology, one that goes beyond localization (i.e., questions such as *where in the brain is syntax processed?*) and beyond passive representation (i.e., questions such as *how is syntax represented in the brain?*). The perspective encourages what we believe are more useful questions with respect to the cognitive architecture (i.e., questions such as *how does the brain support grammatical encoding during speech production?*).

The cognitive architecture of language is about cognition and not about either the products of that cognition (overt language) or the neurobiological machinery that supports it. To understand language use, therefore, we have to focus on performance, not on competence. Chomsky (1965) suggested the reverse, but this simply underlines the point we made earlier—that the explananda of cognitive psychology and linguistics are different. The goal of those working on the cognitive architecture of language should be to understand language performance—because performance (much more directly than competence) is an emergent property of the human brain.

All of this is to say that cognitive psycholinguistics is one of the key components of language science. While we predict that there will be more and more biology in language science as the years go by, we also predict that cognitive psycholinguistics will continue to flourish. It will stand alongside linguistics and neurobiology as a key contributor to solving the puzzle that is human language. This means that language science will increasingly need to be done in multidisciplinary teams, with the cognitive architecture—our understanding of the cognitive underpinnings of linguistic communication—playing an essential part.

REFERENCES

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

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral & Brain Sciences*, *39*, e62.

Cutler, A., & Clifton, Jr., C. (1999). Comprehending spoken language: A blueprint of the listener. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 123–166). Oxford: Oxford University Press.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*(3), 283–321.

Dell, G. S., & Chang, F. (2014). The P-chain: Relating sentence production and its disorders to comprehension and acquisition. *Philosophical Transactions of the Royal Society B-Biological Sciences*, *369*(1634), 20120394.

El Aissati, A., McQueen, J. M., & Cutler, A. (2012). Finding words in a language that allows words without vowels. *Cognition*, *124*(1), 79–84.

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

Evans, N., & Levinson, S. C. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences*, *32*(5), 429–492.

Fodor, J. A. (1983). *Modularity of mind: An essay on faculty psychology*. Cambridge, MA: MIT Press.

Frazier, L., & Fodor, J. D. (1978). The sausage machine: A new two-stage parsing model. *Cognition*, *6*(4), 291–325.

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

Gow, D. W., Segawa, J. A., Ahlfors, S. P., & Lin, F.-H. (2008). Lexical influences on speech perception: a Granger causality analysis of MEG and EEG source estimates. *NeuroImage*, *43*(3), 614–623.

Kingston, J., Levy, J., Rysling, A., & Staub, A. (2016). Eye movement evidence for an immediate Ganong effect. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(12), 1969–1988.

Kittredge, A. K., & Dell, G. S. (2016). Learning to speak by listening: Transfer of phonotactics from perception to production [Special issue]. *Journal of Memory and Language*, *89*, 8–22.

Kleinschmidt, D. F., & Jaeger, T. F. (2015). Robust speech perception: Recognize the familiar, generalize to the similar, and adapt to the novel. *Psychological Review*, *122*(2), 148–203.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*, 1–38.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, *74*(6): 431–461.

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

Marslen-Wilson, W., Tyler, L. K., Waksler, R., & Older, L. (1994). Morphology and meaning in the English mental lexicon. *Psychological Review*, *101*(1), 3–33.

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

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review, 88*(5), 375–407.

McQueen, J. M., Cutler, A., & Norris, D. (2006). Phonological abstraction in the mental lexicon. *Cognitive Science, 30*(6), 1113–1126.

McQueen, J. M., Jesse, A., & Norris, D. (2009). No lexical–prelexical feedback during speech perception or: Is it time to stop playing those Christmas tapes? *Journal of Memory and Language, 61*(1), 1–18.

Meyer, A. S., & Huettig, F. (Eds.). (2016). Speaking and listening: Relationships between language production and comprehension [Special issue]. *Journal of Memory and Language, 89*.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41*(1), 49–100.

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115*(2), 357–395.

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

Norris, D., McQueen, J. M., & Cutler, A. (2016). Prediction, Bayesian inference and feedback in speech recognition. *Language, Cognition and Neuroscience, 31*(1), 4–18.

Page, M. P. A. (2000). Connectionist modelling in psychology: A localist manifesto. *Behavioral and Brain Sciences, 23*(4), 443–467.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 329–347.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology, 7* (pp. 101–139). Berlin: de Gruyter.

Posner, M. I. (2012). *Attention in a social world*. Oxford: Oxford University Press.

Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience, 6(7)*, 576–582.

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96(4)*, 523–568.

Taft, M. (2015). The nature of lexical representation in visual word recognition. In A. Pollatsek & R. Treiman (Eds.), *The Oxford handbook on reading* (pp. 99–113). New York: Oxford University Press.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory and Cognition, 28(4)*, 735–747.