Introducing a Fundamental Problem of Language Science

It is quite difficult to conceptualize how spoken language functions without at once referring to writing. Most readers of the present text began to represent speech with alphabetic signs at about four or five years of age. It is therefore understandable that people with years of training in alphabet writing would view speech as containing combinations of letter-size elements, words, phrases, and sentences, like the units on this page. But it should be recognized that these intuitions are not universal. People who learn writing systems such as Japanese kana or Chinese hànzi characters, for instance, conceptualize units and combinations quite differently, and do not represent letter units, or words as groupings of letters separated by spaces (e.g., Hoosain, 1992; Lin, Anderson, Ku et al., 2011; Packard, 2000).

Nevertheless, specialists in various disciplines have come to use alphabetic signs, along with other orthographic units and categories, not only in analyzing different languages, but also in clinical tests and research on the neural underpinnings of speech processing. For an outsider, this might seem odd. Clearly, not everyone knows alphabet writing. How then could culture-specific concepts of letters and words, or categories like *consonant*, *vowel*, *noun*, *verb* (etc.), serve to analyze different spoken languages, let alone neurobiological processes common to all speakers? Indeed, such applications of orthographic concepts have not gone unchallenged. As documented in this monograph, there is a history of academic work in which authors repeatedly criticize the centrism of analyses that use orthographic units and categories, and many question the face validity of language theories based on such analyses. In considering these theories, one needs to weigh the fact that there is scant evidence – from instrumental observations of speech to brain imaging – validating the view that utterances are processed as sentences containing hierarchical sets of phrases, words, and letter-like units, as conceptualized in alphabet writing and conventional language analysis. Researchers who continue to seek these hierarchical sets in speech or brain responses also face a basic problem in that there are no working definitions of what constitutes a "word," "phrase," or "sentence," except by reference to such marks as spaces in text (Dixon & Aikhenvald, 2002; Haspelmath, 2011). As for the belief that speakers create

sentences using rules that serve to combine given grammatical classes of words, there are conflicting views that are outlined in the present work. For instance, neuroscientists have submitted evidence invalidating the idea that the brain processes words in terms of orthographic-like categories such as *noun* and *verb*, which undermines decades of formal syntactic theory (see, e.g., Vigliocco, Vinson, Druks et al., 2011). Such results and the failure to ground writing-induced concepts of language analysis in speech bear disquieting implications for the field of study. Yet, despite a substantial body of critical commentary, few works present findings that address the problem of the "writing bias" in language science.

For students in psychology, communication disorders, and language science, introductory texts offer little forewarning that core concepts of language study do not link to observable physical or physiological aspects of spoken language. In fact, one pains to find a textbook that mentions the problem. Some works allude to it, though almost as an aside. For example, in the final pages of his *Introduction to Neuropsychology of Spoken Language and Its Disorders*, Ingram (2007) offers the following terse critique of the history of processing models that assume "interim" units of representation such as letter-like phonemes and words:

> Early models of sentence processing (or production) tended to simply borrow the units of interim representation from linguistic theories (competence models). Subsequently, psycholinguists sought evidence from performance constraints for the psychological reality of these units. However, evidence at the neural processing level for an interim level of linguistic representation is scarce at best. (p. 377)

This may have prompted some readers to wonder why such criticism of decades of research appears in the latter pages of a text and not as an introductory warning. Other specialists are just as critical of the prospects that language study might one day link to neuroscience. For example, Poeppel and Embick (2005):

> In principle, the combined study of language and the brain could have effects in several directions. (1) One possibility is that the study of the brain will reveal aspects of the structure of linguistic knowledge. (2) The other possibility is that language can be used to investigate the nature of computation in the brain. In either case, there is a tacit background assumption: namely that the combined investigation promises to generate progress in one of these two domains. Given the actual current state of research, these two positions – rarely questioned or, for that matter, identified in studies of language and the brain – lack any obvious justification when examined carefully. (p. 1)

According to these authors and others (Embick & Poeppel, 2015; Grimaldi, 2012, 2017; Poeppel, 2012), the basic problem is one of *ontological incommensurability*, or the fact that "the fundamental elements of linguistic theory

cannot be reduced or matched up with the fundamental biological units identified by neuroscience" (Grimaldi, 2012, p. 3). Actually, this incommensurability extends beyond neuroscience and includes various domains of instrumental observation. Speech scientists have long recognized there are no divisible units in signals that match letter-like phonemes in language analysis (Liberman, Cooper, Shankweiler et al., 1967). In a recent meta-analysis of neurophysiological models of speech processing, Skipper, Devlin, and Lametti (2017) concluded that "after decades of research, we still do not know how we perceive speech sounds even though this behavior is fundamental to our ability to use language" (p. 78).

However, one notes that the preceding formulation of the issue of ontological incommensurability does not call into question the validity of conventional language analysis, which is not unusual. Poeppel and Embick (2005) assert that, "If asked what to study to learn about the *nature* of language, surely one would not send a student to study neuroscience; rather, one might recommend a course in phonetics or phonology or morphology or syntax or semantics or psycholinguistics" (p. 2). The presumption that linguistic methods serve to analyze the "nature" of language exemplifies a misapprehension of the roots of the ontological problem in language study. It fails to recognize that the issue of incommensurability can arise precisely because units and categories of language analysis do not reflect natural elements, but *cultural* constructs that draw from a writing tradition. At the risk of stating the obvious, students who, for instance, learn to analyze the constituents of "sentences" by examining distributions of symbols in transcripts, or by performing substitutions and commutations of letters, words, parts of words, phrases (etc.), while using such notions as *consonant, vowel, verb, preposition, auxiliary* (etc.), are not working with natural units and categories. They are principally working with orthographic concepts of Latin grammar and overlooking entirely the signals and physiological processes involved in vocal communication. There is, in this sense, a misapprehension in claiming that current methods of language analysis serve to understand the nature of spoken language. As documented in this book, such claims overlook a body of work criticizing the orthographic bias in linguistic methods. In introductory works, though, readers are only marginally informed of the problems that arise when culture-specific concepts of writing are used in analyzing spoken language.

As a further example, in a popular compendium of brain-imaging research on language, Kemmerer (2015b, pp. 273–274) cautions readers against culture-centric assumptions, noting that there are substantial differences in the way languages "carve up" meaning with *words*. Then, a vast body of neuroimaging research on the processing of lexical items is discussed, much of which refers to test protocols involving presentations of

isolated words. Regarding such methods, there are no cautionary remarks that language groups also carve up verbal expressions differently and that meaningful forms may not reflect units like words in European-style writing and dictionaries. This is not simply a technical matter. It is a decisive conceptual issue, one which also carries ethical implications. Although not widely publicized, the cultural specificity of the word concept has led to debates amongst language pathologists confronted by the problem of how to diagnose "word-finding" deficits for speakers of so-called wordless or polysynthetic languages (discussed later on in this monograph). In these cases, linguistic analysis does not serve to distinguish words from phrases or sentences, and the problem is not limited to little known languages like Inuktitut, Mohawk, Cayuga (and others). It can arguably extend to "isolating" languages representing some of the largest communities of speakers, such as Chinese, which does not conceptualize words as in alphabet writing (cf. e.g., Hockett, 1944, p. 255: "there are no words in Chinese"; Packard, 2000, pp. 16 and infra). Certainly, neuroimaging studies that use visual presentations of words like *apple*, *dog*, *cup* (etc.) provide valuable and even essential information on semantic processes and representations. And clearly isolated units corresponding to space-divided words in writing are used to name objects or actions (such as proper names and imperatives). But people do not speak to each other using isolated forms like *apple, dog, cup*, and it is inherently difficult to determine how many "words" there are in basic utterances. For instance, in *I'm done*, *Tom's gone*. *Take'm*, *You're right* (etc.), do *'m, 's, 're*, constitute words? Are the syllables that contain these units subject-verb phrases (and so forth)? Faced with such definitional issues, some analysts contend that, for the most part, speech may not involve words, or combinations of words as in a text, but formulas which are processed as such (Ambridge & Lieven, 2011; Beckner, Ellis, Blythe et al., 2009; Bybee & Beckner, 2009). Thus, beyond cautionary remarks on language relativity, there is the fundamental issue of whether concepts linked to writing can serve to guide research on the structures and processes by which spoken language conveys meaning.

For scientists from other disciplines, the latter concern may seem so basic as to undermine language study as a science. After all, why has the failure to observe notional orthographic units and categories in sensory signals not led to a reevaluation of conventions of language analysis? Some do not see the problem as relating to a writing bias, but instead suggest the need to refine linguistic concepts using "features" and "morphemes" rather than letter-like phonemes and words (Embick & Poeppel, 2015; Poeppel, 2012). Many also point to a general problem of methodology: language specialists need to develop theories that take into account instrumental observations so as to orient what appears to be a stockpiling of eclectic data. For instance, Grimaldi (2012)

notes that there is a basic ontological problem in linking theories of language to observations of neural processes:

> Despite the impressive amount of neural evidence accumulated until now, the field of research results is fragmented and it is quite difficult to reach a unit of analysis and consensus on the object of study. This frustrating state of the art results in a detrimental reductionism consisting in the practice of associating linguistic computation hypothesized at a theoretical level with neurobiological computation. However, these two entities are at the moment ontologically incommensurable. The problem lies in the fact that a theory of language consistent with a range of neurophysiological and neuroimaging techniques of investigation and verifiable through neural data is still lacking. (p. 304)

Grimaldi mentions one exception. In his view, the language theorist Jackendoff has developed a formal model of "sentence" processing that attempts to connect linguistic analyses to neurophysiological observations. The proposal, the *Parallel Architecture Model of Sentence Processing* (Jackendoff, 2007b, 2009, 2017), was principally intended as a response to criticisms that formal theories of sentence generation present static arborescent structures that do not take into account that speech unfolds over time (see Ferreira, 2005). In other words, the model attempted to reconcile an analysis of sentences using static signs on a page with temporal aspects of speech. This is a rather obvious discrepancy between theory and observation, which might lead one to ask why language theories do not generally take into account the time dimension of speech and how this shapes the processing of language. On such issues, Jackendoff's justification for the model is instructive in defining the ontological incommensurability problem that extends well beyond the issue of static formalisms.

In particular, in proposing his model, Jackendoff admitted that he was breaking with a long-standing "mentalist" tradition in language study. By this, he meant that, following Saussure's (1916/1966) division of *langue* and *parole*, language analysts have generally adopted a distinction "between the study of language *competence* – a speaker's f-knowledge ('functional knowledge') of language – and *performance,* the actual processes (viewed computationally or *neurally*) taking place in the mind/brain that put this f-knowledge to use in speaking and understanding sentences" (2009, p. 27). The mentalist perspective adheres to the Saussurean premise that language is a product of the mind that can be studied separately from the speech medium by analyzing transcripts. Although Jackendoff (2009) suggested that the competence–performance division may have originally been intended as a methodological principle (p. 27), this is certainly not the case historically (as clarified in the present monograph). The mentalist tradition, as Jackendoff notes, has been highly influential and was a principal vector of cognitive psychology and cognitive neuroscience. Indeed, G. Miller (2003) recounts that in the 50s,

when behaviorist theories were being overturned by a group of authors for which he was a spokesman, he hesitated to use the term mentalist to describe the views of the group and instead referred to a "cognitivist" approach (p. 142). For Jackendoff, the competence–performance division in the mentalist tradition had some unfortunate consequences in that some theorists "tended to harden the distinction into a firewall: competence theories came to be considered immune to evidence from performance" (2009, p. 28). But this is not the only break with the mentalist tradition that Jackendoff requests for his proposal. To respond to the lack of agreement between formal language theories and neuroscience, he suggests a reassessment of a modular "syntacticocentric" view which has dominated language theory for over half a century and which, in his opinion, was "a scientific mistake" (p. 35).

In the end, the *Parallel Architecture Model* attempts to answer the problem that speech unfolds over time. Even so, the model operates on conventional units and categories where sentences are seen as hierarchical assemblies of letter symbols, words, phrases (etc.), all of which have no general physical or physiological attributes in speech. Again, this incommensurability is not seen as particularly troublesome for the theorist who accepts that units, such as words, "as is well known, are not present in any direct way in the signal" (Jackendoff, 2007a, p. 378), or that letter-like phonemes are not in the "physical world" (Jackendoff, 2017, p. 186). But if such units are not in signals, in physical manifestations of spoken language, then how does neurophysiology operate to extract and process phonemes and word combinations? How do scientists test putative grammars, or how would a child acquire a combinatorial scheme if units are not present in sensory signals? There is in such views a "firewall" of sorts preventing any invalidation of the writing-induced concepts of language analysis and theory. Thus, while some authors acknowledge the historical failure to connect the study of language competence to performance (which includes physiology and signals), the presumption that spoken language can be studied using transcripts and orthographic concepts remains. This has led to enduring pseudo-puzzles that extend to debates on the origins of spoken language. For instance, how would formal grammars based on linguistic analyses emerge in human biology if the units on which the grammars function were not in the physical medium of communication? On such pseudo-puzzles, the failure to ground conventional units of language description in sensory signals has contributed to speculations of their innateness and saltations in the evolution of the mind/brain that present logical problems (as outlined by Christiansen & Chater, 2008).

On the other hand, rather than addressing the problem of using writing concepts in the study of spoken language, some ask if the concepts are not "unavoidable descriptive conveniences" (Bybee & McClelland, 2005; Jackendoff, 2007a, p. 352). Certainly, examining transcripts and orthographic

units can provide useful information in several areas of inquiry, especially in sectors relating to reading. Nonetheless, static writing signs do not represent speech acts and signals. No matter how fine the transcripts are, they do not capture such things as muscle contractions, sound properties, breath flow, or any other physical or physiological aspect of verbal communication. Thus, writing signs offer no means by which to explain the structure of spoken language. As a consequence, researchers who refer to linguistic analyses and theory will not find a definition of the *nature* of units such as letter-like phonemes, words, phrases, or sentences. These forms used to describe spoken language are generally taken as *given* in sectors of language study, but never explained. More importantly, viewing spoken language through writing overlooks the multisensory context of speech that essentially defines the function and meaning of utterances. This latter problem is commonly acknowledged: one cannot interpret the meaning of utterances, let alone understand *how* they convey meaning out of context. Yet current models of language are based on the analysis of script that completely removes the object of study from the motor-sensory medium and communicative environment.

The following monograph presents evidence, some of which has been available for some time, showing that one does not need to refer to writing concepts to study the processes of spoken language. In discussing this evidence, it is a contention that postulates of interim units and categories of language analysis do not exist beyond "descriptive conveniences" for observers who know alphabet writing, and are not "unavoidable" as some have argued. There are observable structures of speech that are universally present across languages. All spoken languages present syllable-like cycles, prosodic groupings, and units of speech breathing. It is also the aim of this book to show that findings linking these structures to neural processes present a major shift in the way semantic representations are conceptualized. Many readers who have formal training in language study are likely to object to such a viewpoint based on the notion that language is separate from speech. In the context where this belief is pervasive, any attempt to address the problem of the incommensurability between conventions of language analysis and observations of speech processes requires a critical look at the historical arguments that have served to maintain a speech–language division. This also extends to a division in methods of inquiry where linguistic analysis is often seen to have theoretical precedence over instrumental observations. These issues have essentially guided the organization of the subject matter of the present monograph into four parts.

Part I, entitled "Questions of Ontology: Writing and the Speech–Language Divide," documents the source of the belief that language is separate from speech. For many historians, this belief originates in the practice, instituted by philologists in the nineteenth century, of viewing spoken language via script.

Influential authors such as Saussure (1857–1913) formulated pivotal arguments for separating *langue* and *parole* that are still echoed in textbooks. In reviewing these arguments, an essential point of Part I is that early language theorists had no instruments by which to record and visualize speech. When instrumental methods became available, the notion that one could work out feature systems by examining letter signs on a page was already accepted in schools of phonology. But a turning point occurred when early instrumentalists reported that writing-induced concepts of linguistic methods did not reflect in speech. At that point, the argument was made that phonological criteria were essential in orienting observations. Many interpreted this to mean that empirical research had to be hypothesis-driven in terms of the assumed units and categories of linguistic theory.

Part II, "Questions of Epistemology: The Role of Instrumental Observations," examines the consequences of the idea of the primacy of linguistic-type analyses in guiding research. It is the case that, in using instrumental techniques, investigators need to have some idea of what exactly in speech serves a communicative function. However, such considerations offer no justification for the belief that research needs to be driven by orthographic concepts of linguistic descriptions. Yet this presumption prevails in sectors of language study. One key example discussed in Part II is the perennial debate on the existence of letter-like phonemes. Despite the acknowledged absence of these units in speech, it has been claimed that types of data, such as spoonerisms transcribed with letters, confirm the existence of phonemes. The assumed primacy of these types of indirect observations over instrumental evidence reflects an epistemological problem and a writing bias that has broad consequences. For instance, some investigators critically refer to studies of spoonerisms in arguing that phonemes are part of an innate competence underlying alphabet writing, a view that has been severely criticized. Part II reviews a body of evidence confronting the belief that language is separate from speech and can be studied using concepts of writing. Of course, not all authors share this viewpoint. Some, in fact, explicitly reject the speech–language division. This entails a major epistemological shift in that abandoning the division implies that structures that are readily identified in motor and sensory aspects of speech *are* the actual structures of spoken language.

Part III, "The Structure of Speech Acts," and Part IV, "The Processing of Speech Meaning," address the problem of the ontological incommensurability between conventional linguistic analysis and observations of spoken language. Part III illustrates how one can study spoken language in context where, instead of focusing on script, one refers to the structural attributes of utterances. Viewed as speech acts, utterances are much like other bodily actions that are performed in a changing environment except that the actions modulate air pressure for communicative purposes. As already mentioned, the structure of

these actions is reflected in syllable-like cycles, temporal chunks, and breath units, all of which emerge from circumscribable processes of motor speech. A central thesis of the present book is that multisensory context information binds to utterance structure, more specifically, to chunks of articulated sounds via mechanisms of motor-sensory coupling and neural entrainment. On this basis, episodic and semantic memory of sensory experiences that accompany utterances can link to action chunks that some authors view as verbal formulas and semantic schemas. The relevance of this approach is discussed in the context of ongoing problems that arise in models of speech and semantic theory that attempt to provide an interface between neural processes and writing-induced concepts of language description.

In sum, Parts I and II offer a review of historical developments underlying a writing bias that has created fundamental problems in research on spoken language, whereas Parts III and IV present an approach that serves to address these problems. However, the background information provided in the first parts should not detract from the main subject of the present monograph. This is not a book about writing, although it refers to critical works on the influence of what some historians call scriptism in language study. It is a book about language science and how one can view processes of spoken language without reference to culture-specific concepts of writing. The latter parts of the monograph develop this view in terms of a body of research findings. This research sheds light on how memory of multisensory experiences binds to structures of motor speech and how activations of semantic or episodic memory in relation to these structures underlies a context-based interpretation of utterances. However, given that many may believe that speech is functionally separate from language and semantics, it is useful to begin by examining how this belief, and conventional concepts of language analysis, arose from a tradition of viewing spoken language through script.